

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 00:05:37 ; Search time 3793 Seconds  
(without alignments)  
2231.705 Million cell updates/sec

Title: US-09-701-674a-23

Perfect score: 179  
Sequence: 1 MGTGSLSDVEDLQEVEMLEEC.....PESDLKEVTVABRLCGTTAS 179

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool.h/US09701674/runat\_15122004.100530.28306/app.query.fasta\_1.327  
-DB=GenEmbl -OPMT=fastap -SUFFIX=Oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 -CGEN 1 1 3731 @runat\_15122004.100530.28306 -NCEU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -USPSLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_cm.\*  
5: gb\_cv.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result , No.	Score	Query Match	Length	DB ID	Description
1	179	100.0	537	9 CR450293	CR450293 Homo sapi
2	179	100.0	1272	6 BD211771	BD211771 Proteins
3	179	100.0	3231	9 BC025697	BC025697 Homo sapi
4	166	92.7	1259	6 CQ718038	CQ718038 Sequence

5	150	83.8	25186	9	AL356109	AL356109 Human DNA
6	120	67.0	1202	10	AF047418	AF047418 Mus muscu
7	120	67.0	1217	10	AF029753	AF029753 Mus muscu
8	120	67.0	1220	10	AB009453	AB009453 Mus muscu
9	120	67.0	1240	10	AF035717	AF035717 Mus muscu
10	120	67.0	1245	10	BC053525	BC053525 Mus muscu
11	120	67.0	1267	10	AF036945	AF036945 Mus muscu
12	118	65.9	1254	9	AF035718	AF035718 Homo sapi
13	109	60.9	187332	2	AC101292	AC101292 Mus muscu
14	109	60.9	241544	2	AC115183	AC115183 Rattus no
15	98	54.7	1164	5	BC933280	BC933280 Gallus ga
16	96	53.6	1257	9	AF047419	AF047419 Homo sapi
17	88	49.2	965	5	BC073597	BC073597 Xenopus l
18	88	49.2	1059	5	AY660871	AY660871 Xenopus l
19	86	48.0	513	10	AF061752	AF061752 Rattus no
20	77	43.0	124070	5	BC530074	BC530074 Zebratish
21	77	43.0	231207	2	CR293516	CR293516 Dario rer
22	67	37.4	189465	2	CR318665	CR318665 Dario rer
23	61	34.1	697	6	AX333347	AX333347 Sequence
24	61	34.1	697	6	AX333763	AX333763 Sequence
25	36	20.1	240	6	AR270465	AR270465 Sequence
26	28	15.6	606	10	AF108216	AF108216 Mus muscu
27	28	15.6	621	9	BT007169	BT007169 Homo sapi
28	28	15.6	621	12	BT008287	BT008287 Synthetic
29	28	15.6	1716	9	AF087036	AF087036 Homo sapi
30	28	15.6	1718	10	AF087035	AF087035 Mus muscu
31	28	15.6	1893	9	AF060154	AF060154 Homo sapi
32	28	15.6	1917	6	CQ717187	CQ717187 Sequence
33	28	15.6	2031	9	BC006313	BC006313 Homo sapi
34	28	15.6	2040	9	BC067827	BC067827 Homo sapi
35	28	15.6	2382	6	AX281674	AX281674 Sequence
36	28	15.6	163698	10	AC115876	AC115876 Mus muscu
37	28	15.6	170668	9	AC104012	AC104012 Homo sapi
38	28	15.6	18157	10	AC121789	AC121789 Mus muscu
39	28	15.6	196331	9	AC028867	AC028867 Homo sapi
40	28	15.6	225399	2	AC113255	AC113255 Rattus no
41	25	14.0	222122	2	CR450806	CR450806 Dario rer
42	21	11.7	107	11	AB094420	AB094420 Sus scro
43	19	10.6	144007	2	AC141743	AC141743 Apis mell
44	18	10.1	65	6	CQ530499	CQ530499 Sequence
45	16	8.9	101	11	BV184299	BV184299 sqnm14420

## ALIGNMENTS

CR450293 537 bp mRNA linear PRI 18-MAY-2004  
Homo sapiens full open reading frame cDNA clone RZPD0834B11D for  
gene TCF21, transcription factor 21; complete cds; without  
stopcodon.

CR450293 1 GI:47496532  
Full ORF shuttle clone, Gateway(TM), complete cds.  
Homo sapiens (human)

CR450293 1 GI:47496532  
Full ORF shuttle clone, Gateway(TM), complete cds.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 537)  
Ebert L., Schick M., Neubert P., Schatten R., Henze S. and Korn B.  
Cloning of human full open reading frames in Gateway(TM) system  
entry vector (pDONR201)

Unpublished  
2 (bases 1 to 537)  
Ebert L., Schick M., Neubert P., Schatten R., Henze S. and Korn B.

Direct Submission  
Submitted (18-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer  
Genomforschung GmbH, Im Neuenheimer Feld 580, D-69120 Heidelberg,  
Germany

CR450293 Homo sapi 76

RZPD; RZPD0834B11D, ORFNO 76  
www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPD0834B11D RZPDLIB;  
Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB NO.  
834

www.rzpd.de/cgi-bin/products/showlib.pl.cgi/response?libNo=834  
 www.rzpd.de/products/orfclones/

Contact: Ina Rolf  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 101  
 Fax: +49 30 32639 111  
 www.rzpd.de

This clone is available from RZPD;  
 contact RZPD (customer.service@rzpd.de) for further information.  
 This CDS clone is a part of a collection of human full length  
 expression clones generated by RZPD.

This CDS has been cloned without stopcodon.  
 The CDS has been inserted into pDONR201 via a BP Clonase(TM)  
 reaction. Additional sequence has been added in front of the start  
 codon (ATG): att. AAAAAA GGT GGC ACC CCT GGT CCA GGT (ATG)  
 After the last codon additional sequence has been added: CCA GGC  
 CCA GGC GGC G in front of the 3' att site (AC CCA GCT TTC TT).  
 Compared to the reference sequence NM\_003206 we did not find any  
 amino acid exchanges.

Clone distribution: <http://www.rzpd.de/products/orfclones/>.

#### FEATURES

source  
 1..537  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="RZPD08348111D"  
 /clone\_lib="Human Full ORF Clones Gateway(TM) - RZPD"  
 /lab\_host="DH10B"  
 /note="Vector: pDONR201, Site\_1: attP1; Site\_2: attP2"  
 1..537  
 /genes="TCF21"  
 1..537  
 /genes="TCF21"  
 /codon\_start=1  
 /protein\_id="CAG29289.1"  
 /db\_xref="GI:7496533"  
 /translation="MSGSLSDVELQVEMLECDGLKWDNSKEFTVNSTERSNC  
 ENSPKRGGLGKRRKPTKKSPGVSQEGKQVORNAARERARVRVLSKAFSL  
 KTLTPWPPDTKLSDTLRLASSYIAHLRLANDKYENGYIHPVNLTPFWAGKP  
 ESDLKEVVTASRLCGTTS"

#### ORIGIN

Alignment Scores:  
 Pred. No.: 2,71e-184 Length: 537  
 Score: 179.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x CR450293 (1-537)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 1 ATGTCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGAGATGTGGAATGT 60  
 Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 61 GACGGGTGTGAATGATTCGAACAGGAATTGTGACTTCCACGAGACGACCGAGGAG 120  
 Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60  
 Db 121 AGCTCAACTCGCAGAAATGGTCTCCCAAGAGGCGCGCGCTGGCGCAAGAGGAGG 180  
 Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysValGln 80  
 Db 181 AAGGCCCCCAACCAAGAGAGCCCTTGGAGCGGGTTCAGCAGAGGGAAGCAGGTCCAG 240  
 Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 241 CGCAACGCCGCCACACCGCGGAGGCGGGCCCGCATCGAGATGCTGAGCAAGCCCTTCC 300

Qy 101 ArgLeuLysThrThrLeuProTtpValProAspThrLysLeuSerLysLeuAspThr 120  
 Db 301 AGATCAGACACCCCTGCTGGTGGTCCCGCCGACACCAAGCTCTCCAGCTGGACAG 360  
 Qy 121 LeuArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 361 CTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCCTGGCTAACGACAAATAC 420  
 Qy 141 GluAsnGlyTyrlleHisProValAsnLeuThrTtpProPheMetValAlaGlyLysPro 160  
 Db 421 GAGAACGGGTACATTCACCGGTCACCTGACGTGGCCCTTATGTGGCCGCGGAACCC 480  
 Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 481 GAGAGTGACCTGAAGAAGTGGTGACCGGAGCGGCTTATGTGAACCAACCGCGGTCC 537

#### RESULT 2

LOCUS BD211771 1272 bp DNA linear PAT 17-JUL-2003  
 DEFINITION Proteins regulating gene expression.  
 ACCESSION BD211771  
 VERSION BD211771.1 GI:33021541  
 KEYWORDS JP 2002517246-A/23  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1272)  
 AUTHORS Lal, P., Yue, H., Tang, Y.T., Hillman, J.L., Bandman, O., Corley, N.C.,  
 Guegler, X.J., Gorgone, G.A., Baughn, M.R., Patterson, C. and Lu, D.A.M.  
 TITLE Proteins regulating gene expression  
 JOURNAL Patent: JP 2002517246-A 23 18-JUN-2002;  
 COMMENT OS Homo sapiens (human)  
 INCYTE PHARMACEUTICALS INC  
 PN JP 2002517246-A/23  
 PD 18-JUN-2002  
 PF 11-JUN-1999 JP 2000553586  
 PR 12-JUN-1998 US 60/089029, 29-JUL-1998 US 60/094575 PR  
 14-OCT-1998 US 60/104624  
 PI PREETI LAL, HENRY YUE, Y TOM TANG, JENNIFER L HILLMAN, OIGA PI  
 BANDMAN,  
 PI NEIL C CORLEY, KARL J GUEGLER, GINA A GORGONE, MARIAH R BAUGHN,  
 PI CHANDRA PATTERSON, DYUNG AINA M LU  
 PC C12N15/00, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61P15/00,  
 PC A61P25/00,  
 PC A61P35/00, A61P37/02, A61P43/00, C07K14/47, C07K16/18, C12N15/00,  
 PC A61K37/02  
 CC Incyte Clone No.: 3340296  
 FH Key Location/Qualifiers  
 FT source 1..1272  
 FT /organism="Homo sapiens (human)".

#### FEATURES

source  
 1..1272  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

#### ORIGIN

Alignment Scores:  
 Pred. No.: 6.16e-184 Length: 1272  
 Score: 179.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x BD211771 (1-1272)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 268 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGAGATGTGGAATGT 327  
 Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db 328 GACGGGTGAATGATTTCGACCAAGGAATTTGTGACTTCCACGAGACCCAGGAG 387  
 Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 Db 388 AGCTCCAACTGCGAGATGGGTCTCCCGACAGAGGCGCGGGGCTGGCGAAGAGGAG 447  
 Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 448 AAGCGCCCCACCAAGAGAGCCCTGAGCGGGGTGAGCCAGAGGGGAGAGGTCCAG 507  
 Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 508 CGCAGACGCCGCCAACCGGAGCGGCGCGCATGCGAGTCTGACGAGGAGCTTCTCC 567  
 Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 568 AGACTCAAGACCAACCCCTGCGTGGTGGCCCGCCGACCAAGCTCTCCAAGCTGGACAG 627  
 Qy 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 628 CTGAGGCTGGCTGCGTACATCGCCCACTTGAAGCAGATCTGCTGACGAGGAGCTTCTCC 687  
 Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 688 GAGACGGGTACATTCACCGGCTCAACCTGACGTGGCCCTTTATGGTGGCCGGAAACCC 747  
 Qy 161 GluSerAspLeuLysGluValThrAlaSerArgLysCysGlyThrThrAlaSer 179  
 Db 748 GAGAGTGACCTGAAGAAGTGGTGACCGGAGCGGCTTATGTGGAACCAACCGCGTCC 804

## RESULT 3

BC025697

LOCUS

DEFINITION

Homo sapiens transcription factor 21, transcript variant 1, mRNA

(cdna clone MGC:34534 IMAGE:5221575), complete cds.

ACCESSION

BC025697

VERSION

MGC.

KEYWORDS

MGC.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 3231)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, J., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Cabavant, T.L., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

REFERENCE

2 (bases 1 to 3231)

AUTHORS

Strausberg, R.

Direct Submission

Submitted (06-MAR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

JOURNAL

PUBMED

TITLE

REFERENCE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

JOURNAL

TITLE

## REMARK

COMMENT

NIH-MGC Project URL: http://mgs.nci.nih.gov

Contact: MGC help desk

Email: cgaaps-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: http://www.nisc.nih.gov/

Contact: nisc.mgc@nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghigati, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,

Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IRAK Plate: 49 Row: k Column: 7

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 4507394.

## FEATURES

Location/Qualifiers

1..3231

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="MGC:34534 IMAGE:5221575"

/tissue="types=Pancreas, Spleen, adult pooled"

/clone\_lib="NIH MGC\_120"

/lab\_host="DH10B"

/note="vector: pCMV-SPORT6"

1..3231

gene

/gene="TCF21"

/note="synonym: POD1"

/db\_xref="LocusID:6943"

/db\_xref="MIM:603306"

260..799

CDS

/gene="TCF21"

/codon\_start=1

/product="transcription factor 21"

/protein\_id="AAH25697.1"

/db\_xref="GI:19344016"

/db\_xref="LocusID:6943"

/db\_xref="MIM:603306"

/translation="MSTGSLSDVEDLQEVEMLECDGLKMDSNKEFVTSNESSNC

ENGSPQKRGGLGKRRKAPTKPSLGSQEGKQVQRNANARERARMRLSKAFSL

KTPLPWPPTKLKSLDITLGLASSYIAHLRQILANDKYENGYHPVNLTPFWVAGKP

ESDLKEVVTASRLCGTAS"

ORIGIN

Alignment Scores:

Pred. No.: 1.5e-183

Score: 179.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 100.00%

DB: 9

Length: 3231

Matches: 179

Conservative: 0

Mismatch: 0

Indels: 0

Gaps: 0

US-09-701-674A-23 (1-179) x BC025697 (1-3231)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20

Db 260 ATGTCACCGGCTCCCTCAGGATGTGAGGACCTTCAGAGGTGGAGATGTTGAATGT 319

Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40

Db 320 GACGGGTGAAATGGATTCGACAGGAATTTGTGACTTCCACGAGACCCAGGAG 379

Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60

```

Db      380 AGTCTCACTGCGAGAAATGGTCTCCAGAGAGGGCGGGCGGCTGGCGAGAGGAGG 439
Qy      61  LyAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      440 AAGCGCGCCACCAAGAAGAGCCCTCAGCGGGGTAGCCAGAGGGGAGAGGAGGTCCAG 499
Qy      81  ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      500 CGCAAGCCGCCAACCGGAGAGCGGCCCGCATCGAGTGTCTGAGCAAGGCCCTTCTC 559
Qy      101 ArgLeuLysThrThrLeuProTrrpValProProAspThrLysLeuSerLysLeuAspThr 120
Db      560 AGACTCAAGACCAACCCCTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 619
Qy      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleAlaAsnAspLysTyr 140
Db      620 CTGAGGTGGCGGTCAGCTACATCGGCCCACTTGAGGCAGATCCTGGCTAACGACAAATAC 679
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db      680 GAGACGGGTACATTCACCGGTCACCTGACGTGGCCCTTTATGCTGGCGCGGAAACCC 739
Qy      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      740 GAGAGTGACCTGAAGAAGTGTGTGACCGAGCGGCTTATGTGGAACCAACCGCGTCC 796

RESULT 4
LOCUS      CQ718038                1259 bp    DNA        linear    PAT 03-FEB-2004
DEFINITION Sequence 3972 from Patent WO02068579.
ACCESSION  CQ718038
VERSION    CQ718038.1  GI:42278895
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS   Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
JOURNAL   Patent: WO 02068579-A 3972 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES   Location/Qualifiers
            source
            1..1259
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      7,76e-170      Length:      1259
Score:          166.00         Matches:    179
Percent Similarity: 99.44%     Conservative: 0
Best Local Similarity: 99.44%  Mismatches: 0
Query Match:     92.74%       Indels:    1
DB:              6            Gaps:      0

US-09-701-674A-23 (1-179) x CQ718038 (1-1259)
Qy      1  MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db      264 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAGAGGTGAGATGTGGAAATGT 323
Qy      21  AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      324 GACGGGTGAAAATGGATTGCAACAAAGGAATTTGTGACTTCCAAACGAGAGCACCGAGGAG 383
Qy      41  SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db      384 AGCTCCAACTGCAGAAATGGTCTCCCAAGAGCGCGCGGCTGGCGCAAGAGGAGG 443

```

```

Qy      61  LyAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      444 AAGCGCGCCACCAAGAAGAGCCCTCAGCGGGGTAGCCAGAGGGGAGAGGAGGTCCAG 503
Qy      81  ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSe 100
Db      504 CGCAACGCCGCCAACCGGAGAGCGGGGCCCGCATCGAGTGTCTGAGCAAGGCCCTTCTC 563
Qy      100 rArgLeuLysThrThrLeuProTrrpValProProAspThrLysLeuSerLysLeuAspTh 120
Db      564 CAGACTCAAGACCAACCCCTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 623
Qy      120 rLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleAlaAsnAspLysTyr 140
Db      624 GCTCAGGCTGGCGTCCAGCTACATCGGCCCACTTGAGGCAGATCCTGGCTAACGACAAATA 683
Qy      140 rGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPr 160
Db      684 CGAAGACGGGTACATTCACCGGTCACCTGACGTGGCCCTTTATGCTGGCGCGGAAACC 743
Qy      160 oGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      744 CGAGAGTGACCTGAAGAAGTGTGTGACCGGCGGCGCTTATGTGGAACCAACCGCGTCC 801

AL356109                25186 bp    DNA        linear    PRI 23-JAN-2001
LOCUS      Human DNA sequence from clone RP11-373A10 on chromosome 6. Contains
DEFINITION ESTs, STSs, GSSs and CpG islands. Contains the TCF21 gene for
ACCESSION  AL356109
VERSION    AL356109.8  GI:10443434
KEYWORDS   HTG; CpG island; TCF21; transcription factor 21.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 25186)
AUTHORS   Wallis, J.
TITLE     Direct Submission
JOURNAL   Submitted (15-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT    On Oct 1, 2000 this sequence version replaced gi:10186519.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/Hgp/Chr6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-373A10 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP1-73H22 is at 25087 in this sequence.
            The true right end of clone RP4-662A9 is at 100 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the 'unsure' feature key
            is ambiguous, there is an annotation using the 'unsure' feature key.
            RP11-373A10 is from the library RPI1-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm

```





/note="MLT1H repeat: matches 1. .286 of consensus"  
 24559. 25036  
 /note="MUT1D repeat: matches 1. .505 of consensus"  
 complement (24819. .25186)  
 /note="match: STS: Em:G30455"

## repeat\_region

## misc\_feature

## ORIGIN

## Alignment Scores:

Pred. No.: 3 08e-151 Length: 25186  
 Score: 150.00 Matches: 150  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 83.80% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x AL356109 (1-25186)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 13182 ATGTCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGAATGT 13241  
 Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 13242 GACGGTGTGAATGATTCGAACAGGAATTTGTGACTTCCACGAGACCCAGGAG 13301  
 Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
 Db 13302 AGCTCCAACTCGAGAATGGTCTCCCAAGAGGCGCGGGCGCTGGCAAGAGGAGG 13361  
 Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluLysGlnValGln 80  
 Db 13362 AAGCGGCCCAACAGAGAGCCCTTGGCGGGTCCAGCAGAGGAGGAGGAGGTCACG 13421  
 Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 13422 CGCAACGCCGCCAACCGGAGAGCGGCCCGCATCGAGTGTGAGCAAGGCCCTTCTCC 13481  
 Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 13482 AGACTCAAGACCACTCCCTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 13541  
 Qy 121 LeuArgLeuAlaSerSerTrilleAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140  
 Db 13542 CTGAGCTGGCGTCCAGCTACATCGCCACTTGGAGCAGATCTTGCTAACGCAATAC 13601  
 Qy 141 GluAsnGlyTyrIleHisProValAsnLeu 150  
 Db 13602 GAGACGGGTACATTCAACCGGTCAACCTG 13631

## RESULT 6

## LOCUS

AF047418 Mus musculus epicardin mRNA, complete cds. linear ROD 06-OCT-1998

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

1 (bases 1 to 1202)  
 Robb,L., Mifsud,L., Hartley,L., Biben,C., Copeland,N.G.,  
 Gilbert,D.J., Jenkins,N.A. and Harvey,R.P.  
 epicardin: A novel basic helix-loop-helix transcription factor gene  
 expressed in epicardium, branchial arch myoblasts, and mesenchyme  
 of developing lung, gut, kidney, and gonads  
 Dev. Dyn. 213 (1), 105-113 (1998)  
 9733105  
 Robb,L. and Mifsud,L.  
 Direct submission  
 Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza

## FEATURES

## source

1. .1202

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="Balb/c"

/db\_xref="taxon:10090"

199. .738

/function="basic helix-loop-helix transcription factor"  
 /notes="expressed in embryonic epicardium and  
 mesenchyme-derived tissues of lung, gut, kidney, and  
 gonad"

/codon\_start=1

/product="epicardin"

/protein\_id="AAC62531.1"

/db\_xref="GI:2921851"

/translation="MSTGSLSDVEDIQEVMLDCDSLKVDNKEFCTSNSTEGSNC  
 ENSGPKRGGLGRRKAPTKGSLGSGQEGKQVORNANARERMRVLSKAFSL  
 KTTLPWPPDTKSLDTLFLASSYLAHLRQILANDKYENGYIHPVNLTPFWFVAGKP  
 ENLKEYVTANRLCGTTAS"

## ORIGIN

## Alignment Scores:

Pred. No.: 5.99e-120 Length: 1202  
 Score: 120.00 Matches: 120  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 67.04% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF047418 (1-1202)

Qy 42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArgLys 61  
 Db 322 TCCAACATGCGAGAACGGGTCTCCACAGAGGGTGGCGGTGGCTAGGCAAGAGGAGGAG 381  
 Qy 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81  
 Db 382 GCGCCCACTAAGAAAACCCCGCTCAGCGGGTCCAGCAGAGGGCAAGAGTCCAGCGC 441  
 Qy 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101  
 Db 442 AACCGGCCAATGCTCGTGGCGGCCCGCATCGGGTGTGAGCAAGGCCCTTCTCCAGG 501  
 Qy 102 LeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThrLeu 121  
 Db 502 CTCGAAGACCACTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 561  
 Qy 122 ArgLeuAlaSerSerTrilleAlaHisLeuArgGlnLeuAlaAsnAspLysTyrGlu 141  
 Db 562 AGGTGGCGTCCAGTACATCGCTCACTTAAGCAGATCTTGGCCACGACAGTACGAG 621  
 Qy 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161  
 Db 622 AACGGTTACATTCAACCACTCACTGACGTGGCCCTTTATGTGGCGGCAACACGAG 681

## RESULT 7

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

AF029753 Mus musculus basic helix-loop-helix factor Cori mRNA, complete cds.  
 AF029753.1  
 AF029753.1 GI:2613018  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1217)  
 Hidai,H., Bardales,R., Goodwin,R., Quettermous,T. and  
 Quettermous,E.E.  
 Cloning of capulin, a basic helix-loop-helix factor expressed in  
 progenitor cells of the pericardium and the coronary arteries

JOURNAL Mech. Dev. 73 (1), 33-43 (1998) April 102(a)

MEDLINE 98213629

PUBMED 9545226

REFERENCE 2 (bases 1 to 1217)

AUTHORS Quettermous, T., Quettermous, E.E., Hida, H., Padel, B., Boutet, S.C. and Bardales, R.

TITLE Direct Submission

JOURNAL Submitted (10-OCT-1997) Department of Medicine, Division of Cardiology, Stanford University, Palk Bldg, 300 Pasteur Drive, Stanford, CA 94305, USA

FEATURES Location/Qualifiers

source 1..1217

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/tissue\_type="pericardium and coronary arteries"

gene 1..1217

/genes="Cor1"

CDS 213..752

/gene="Cor1"

/function="transcription factor"

/codon\_start=1

/product="basic helix-loop-helix factor Cor1"

/protein\_id="AB84256.1"

/db\_xref="GI:2613019"

/translation="MSTGSLSDVEDLQEVMLDCDSLVKDSNKEFGTNSNTEGSGNC ENGSPQGRGLGKRRKAPT KKSPLSGVQEGKQVORNAANARERMRVLKAFSRL KTLTPWVPPTKLSKLDLRLASSYIAHLRQLANDKYENGVIHPVNLTPWPMVAGKP ENDLKEVVTANRLCGTTAS"

## ORIGIN

Alignment Scores:

Pred. No.:	6.06e-120	Length:	1217
Score:	120.00	Matches:	120
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	67.04%	Indels:	0
DB:	10	Gaps:	0

US-09-701-674A-23 (1-179) x AF029753 (1-1217)

QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyValArgGlyLeuGlyLysArgArgLys 61

DB 336 TCCAACTGCGAGAACGGGTCTCCACAGAGGGTGGCGTGGCTAGGAGGAGGAGGAG 395

QY 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81

DB 396 GCSCCCCACTAAGAAAAAGCCGCTCAGCGGGGTCCAGCAGAGGGCAAGCAGGTCCAGCGC 455

QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101

DB 456 AACGCGGCCCATGCTGTGAGCGGGCCCGCATGCGGGTGTCTGAGCAAGGCCCTTCTCCAGG 515

QY 102 LeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThrLeu 121

DB 516 CTCAGACCAACCTGCGCTGGGTGGCCCGGACACCAAGCTCTCCAAAGTGGACACTCTC 575

QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141

DB 576 AGCTGGCGTCCAGCTACATCGCTCACTTAAGCAGATCCTGGCCCAACCAAGTACGAG 635

QY 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161

DB 636 AACGGTTACATTACCCAGTCACTGACGTGGGCCCTTTATGGTGGCCGCGCAACACGAG 695

RESULT 8

AB009453

LOCUS Mus musculus mRNA for nephogonadin, complete cds. ROD 11-DEC-1997

DEFINITION

AB009453.1

ACCESSION

AB009453.1

VERSION

AB009453.1

KEYWORDS

nephogonadin. (house mouse)

SOURCE

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1220)

AUTHORS Tamura, M. and Nakatsuji, N.

TITLE Identification of Nephogonadin, a novel basic helix-loop-helix gene Published Only in Database (1997)

REFERENCE 2 (bases 1 to 1220)

AUTHORS Tamura, M. and Nakatsuji, N.

TITLE Direct Submission

JOURNAL Submitted (04-DEC-1997) Masaru Tamura, National Institute of Genetics, Mammalian Development; Yata 1111, Mishima, Shizuoka 411, Japan (E-mail:matamura@lab.nig.ac.jp, Tel:+81-559-81-6832, Fax:+81-559-81-6828)

FEATURES Location/Qualifiers

source 1..1220

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

223..762

/function="transcription factor"

/note="basic helix-loop-helix gene"

/codon\_start=1

/product="nephogonadin"

/protein\_id="BAA23883.1"

/db\_xref="GI:2696118"

/translation="MSTGSLSDVEDLQEVMLDCDSLVKDSNKEFGTNSNTEGSGNC ENGSPQGRGLGKRRKAPT KKSPLSGVQEGKQVORNAANARERMRVLKAFSRL KTLTPWVPPTKLSKLDLRLASSYIAHLRQLANDKYENGVIHPVNLTPWPMVAGKP ENDLKEVVTANRLCGTTAS"

polyA\_site 1220

/note="17 A nucleotides"

ORIGIN

Alignment Scores:

Pred. No.:	6.07e-120	Length:	1220
Score:	120.00	Matches:	120
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	67.04%	Indels:	0
DB:	10	Gaps:	0

US-09-701-674A-23 (1-179) x AB009453 (1-1220)

QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyValArgGlyLeuGlyLysArgArgLys 61

DB 346 TCCAACTGCGAGAACGGGTCTCCACAGAGGGTCCCGTGGCTAGGAGGAGGAGGAG 405

QY 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGlyLysGlnValGlnArg 81

DB 406 GCGCCCACTAAGAAAAAGCCGCTCAGCGGGTCCAGCAGAGGGCAAGCAGGTCCAGCGC 465

QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101

DB 466 AACGCGCCAATGCTGTGAGCGGGCCCGCATGCGGGTGTCTGAGCAAGGCCCTTCTCCAGG 525

QY 102 LeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThrLeu 121

DB 526 CTCAGACCAACCTGCGCTGGGTGGCCCGGACACCAAGCTCTCCAAAGTGGACACTCTC 585

QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141

DB 586 AGGTGGCGTCCAGCTACATCGCTCACTTAAGCAGATCCTGGCCCAACCAAGTACGAG 645

QY 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161

DB 646 AACGGTTACATTACCCAGTCACTGACGTGGGCCCTTTATGGTGGCCGCGCAACACGAG 705

RESULT 9

AF035717

LOCUS Mus musculus mesoderm-specific basic-helix-loop-helix protein

DEFINITION

AF035717

VERSION

AF035717

KEYWORDS

mus musculus mesoderm-specific basic-helix-loop-helix protein

SOURCE

```

ACCESSION AF035717
VERSION AF035717.1 GI:2745884
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.
TITLE Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed in mesenchymal and glomerular epithelial cells in the developing kidney
JOURNAL Mech. Dev. 71 (1-2), 37-48 (1998) Feb 16(2a)
MEDLINE 98175875
PUBMED 9507058
REFERENCE 2 (bases 1 to 1240)
AUTHORS Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1997) Internal Medicine, Yale University, 333 Cedar Street, New Haven, CT 06520-8029, USA
FEATURES
    source
    1. 1240
        /organism="Mus musculus"
        /mol_type="mRNA"
        /db_xref="taxon:10090"
        /chromosome="10"
        /map="near D10Mit2"
    gene
    1. 1240
        /genes="Pod1"
        /gene="Pod1"
        /note="Pod-1"
        /codon_start=1
        /product="mesoderm-specific basic-helix-loop-helix protein"
        /protein_id="AAC62513.1"
        /db_xref="GI:2745885"
        /translation="MSGTSLSDVDELDQVEMDCSLKVDNKFQTSNESTEEGNC
        ENGSPKGGGLGKPKRKATKPSLGVSGVSGEKGQVORNAARERARVLSKAPSL
        KTLPLWPDPDKLSKLDTLRLASSVIAHLRLQILANDKYENGVIHPVNLTFWPFVWAGKP
        ENDLKSVTVNLRGCTTAS"
    misc_feature
    457..618
        /genes="Pod1"
        /note="encodes basic-helix-loop-helix domain"

ORIGIN
Alignment Scores:
Pred. No.: 6,17e-120 Length: 1240
Score: 120.00 Matches: 120
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.04% Indels: 0
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF035717 (1-1240)
QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgLys 61
Db 349 TCCAACTGCGAGAACGGGCTCCACAGAGAGGTCGGGTGGCTAGGCAAGAGGAGGAG 408
QY 62 AlaProThrLysSerProLeuSerGlyValSerGlnGluGlyValGlnArg 81
Db 409 GCGCCCACTAAGAAACCCCGCTCAGCGGGGTGACGAGAGGGAAGCAGGTCACGGC 468
QY 82 AsnAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
Db 469 AACCGCGGCAATGCTGTCGGGGGGCCGATCGGGGTGCTGAGCAAGGCTTCCAGG 528
QY 102 LeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThrLeu 121
Db 529 CTCAGACACACCCCTGGCTGGTGGTGGCCCGGACCAAGCTCTCAAGGTGACACTCTC 588
QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141

```

```

Db 589 AGGCTGGCGTCCAGCTACATCGTCACTTAAGGCAGATCTCGGCCAACGCAAGTAGCAG 648
QY 142 AsnGlyTyrIleHisProValAsnLeuThrTrpPheMetValAlaGlyLysProGlu 161
Db 649 AACGGTTACATTACCAGTCAACCTGACGTGACGTCCTTTATGGTGGCCGCAACAGGAG 708

RESULT 10
LOCUS BC053525
DEFINITION BC053525 1246 bp mRNA linear ROD 30-JUN-2004
IMAGE:6529241, complete cds.
ACCESSION BC053525
VERSION BC053525.1 GI:31566109
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1246)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1246)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Ioshizuka, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalys, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1246)
Strausberg, R.
Direct Submission
Submitted (09-JUN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gqabs-x@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.ncn@nih.gov
Ahter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 108 Row: p Column: 16.
Location/Qualifiers

```

```

source
1. .1246
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVE/N"
/db_xref="taxon:10090"
/clone="MGC:58914 IMAGE:6529241"
/tissue_type="Colon, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_C024"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .1246
/genes="Tcf21"
/note="synonyms: epc, Pod1, capsulin, epicardin, Pod-1"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
233. .772
/genes="Tcf21"
/codon_start=1
/product="transcription factor 21"
/protein_id="AAH53525.1"
/db_xref="GI:31566110"
/db_xref="LocusID:21412"
/db_xref="MGI:1202715"
/translation="MSTGSLSDVEDLQEVEMLDCLSKVDSNKEFGTSTEEGNSC
ENGSPKGRGLGKRKAPTCKPSLGSQVQKQVORNAANARERARMRVLSKAFSRL
KTTLPVPDPTKLSKLDLRLASSYIAHLRLQILANDKYENGYIHPVNLTPFWVAGKP
ENDLKEVVTANRLCGTTAS"
ORIGIN
Alignment Scores:
Pred. No.: 6.2e-120 Length: 1246
Score: 120.00 Matches: 120
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.04% Indels: 0
DB: 10 Gaps: 0
US-09-701-674A-23 (1-179) x BC053525 (1-1246)
QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyValSerGlyLeuGlyLysArgArgLys 61
DB 356 TCCAACTGGAGAACCGGTCTCCACAGAGGGTGGCGGTGGCTAGGCAAGAGGAGAG 415
QY 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81
DB 416 GCGCCCACTAAGAAAAGCGCTCAGCGGGTCAGCCAGGAGGCAAGCAGGTCCAGCGC 475
QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
DB 476 AACCGCGCAATGCTCGTGGCGGGCCGCGCATGCGGGTCTGAGCAAGGCCCTCTCCAGG 535
QY 102 LeuLysThrThrLeuProTyrValProProAspThrLysLysLysLeuAspThrLeu 121
DB 536 CTCACAGCACCCCTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 595
QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141
DB 596 AGGCTGGCTCCAGCTACATCGCTCACTTAAGCAGATCTCTGCGCAACAGACAGTACGAG 655
QY 142 AsnGlyTyrIleHisProValLeuLeuThrTyrProPheMetValAlaGlyLysProGlu 161
DB 656 AACGGTTACATTCCACCGATCAACCTGACGTGGCCCTTTATGGTGGCGGCAACAGAG 715
RESULT 11
AF036945 1267 bp mRNA linear ROD 15-JUN-1998
LOCUS Mus musculus capsulin mRNA, complete cds.
DEFINITION AF036945
ACCESSION AF036945.1 GI:2695694
VERSION AF036945.1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 1267)
Lu,J., Richardson,J.A. and Olson,E.N.
Capsulin: a novel bHLH transcription factor expressed in epicardial
progenitors and mesenchyme of visceral organs
Mech. Dev. 73 (1), 23-32 (1998) Apr 002 (a)
98213628
PUBMED
MEDLINE
REFERENCE 2 (Bases 1 to 1267)
AUTHORS Lu,J. and Olson,E.N.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1997) Molecular Biology and Oncology, UT
Southwestern Medical Center, 6000 Harry Hines Blvd., NA8.510,
Dallas, TX 75235-9148, USA
FEATURES
source
1. .1267
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
264. .803
/function="BHLH transcription factor"
/codon_start=1
/product="capsulin"
/protein_id="AAC23537.1"
/db_xref="GI:2695695"
/translation="MSTGSLSDVEDLQEVEMLDCLSKVDSNKEFGTSTEEGNSC
ENGSPKGRGLGKRKAPTCKPSLGSQVQKQVORNAANARERARMRVLSKAFSRL
KTTLPVPDPTKLSKLDLRLASSYIAHLRLQILANDKYENGYIHPVNLTPFWVAGRP
ENDLKEVVTANRLCGTTAS"
ORIGIN
Alignment Scores:
Pred. No.: 6.3e-120 Length: 1267
Score: 120.00 Matches: 120
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.04% Indels: 0
DB: 10 Gaps: 0
US-09-701-674A-23 (1-179) x AF036945 (1-1267)
QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyValSerGlyLysArgArgLys 61
DB 387 TCCAACTGGAGAACCGGTCTCCACAGAGGGTGGCGGTGGCTAGGCAAGAGGAGAG 446
QY 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81
DB 447 GCGCCCACTAAGAAAAGCGCTCAGCGGGTCTAGCCAGGAGGCAAGGTCACGCGC 506
QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
DB 507 AACCGCGCAATGCTCGTGGCGGGCCGCGCATGCGGGTCTGAGCAAGGCCCTCTCCAGG 566
QY 102 LeuLysThrThrLeuProTyrValProProAspThrLysLysLysLeuAspThrLeu 121
DB 567 CTCACAGCACCCCTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 626
QY 122 ArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141
DB 627 AGGCTGGCTCCAGCTACATCGCTCACTTAAGCAGATCTCTGCGCAACAGACAGTACGAG 686
QY 142 AsnGlyTyrIleHisProValLeuLeuThrTyrProPheMetValAlaGlyLysProGlu 161
DB 687 AACGGTTACATTCCACCGATCAACCTGACGTGGCCCTTTATGGTGGCGGCAACAGAG 746
RESULT 12
AF035718 1254 bp mRNA linear PRI 05-OCT-1998
LOCUS Homo sapiens mesoderm-specific basic-helix-loop-helix protein
DEFINITION AF035718
ACCESSION AF035718
VERSION AF035718.1 GI:2745886
KEYWORDS

```



Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

# TITLE JOURNAL

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:17060067.

## COMMENT

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16369

Center Clone name: 103.E.4

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178139 bases at least Q40

Consensus quality: 182376 bases at least Q30

Consensus quality: 183759 bases at least Q20

Insert size: 195000; agarose-fp

Insert size: 184732; sum-of-contigs

Quality coverage: 6.4 in Q20 bases; agarose-fp

Quality coverage: 6.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 23755: contig of 23755 bp in length  
\* 23756: gap of 100 bp  
\* 23856: contig of 674 bp in length  
\* 24529: gap of 100 bp  
\* 24630: contig of 1166 bp in length  
\* 25795: gap of 100 bp  
\* 25796: gap of 100 bp  
\* 25996: contig of 1079 bp in length  
\* 26974: gap of 100 bp  
\* 26975: gap of 100 bp  
\* 27074: gap of 100 bp  
\* 27075: contig of 1530 bp in length  
\* 28604: gap of 100 bp  
\* 28705: contig of 1363 bp in length  
\* 30068: gap of 100 bp  
\* 30168: contig of 921 bp in length  
\* 31089: gap of 100 bp  
\* 31189: contig of 1770 bp in length  
\* 32958: gap of 100 bp  
\* 33059: contig of 2172 bp in length  
\* 35331: gap of 100 bp  
\* 37407: contig of 2077 bp in length  
\* 37508: gap of 100 bp  
\* 40202: contig of 2695 bp in length  
\* 40302: gap of 100 bp  
\* 40303: contig of 3531 bp in length  
\* 43934: gap of 100 bp  
\* 46356: contig of 2423 bp in length  
\* 46457: gap of 100 bp  
\* 48641: contig of 2185 bp in length  
\* 48741: gap of 100 bp  
\* 48742: contig of 2939 bp in length  
\* 51681: gap of 100 bp  
\* 51781: contig of 3211 bp in length  
\* 55091: gap of 100 bp  
\* 55092: contig of 3487 bp in length  
\* 58578: gap of 100 bp  
\* 58579: contig of 3506 bp in length  
\* 62184: contig of 3506 bp in length

Alignment Scores:

\* 62185 62284: gap of 100 bp  
\* 62285 69983: contig of 7699 bp in length  
\* 69984 70083: gap of 100 bp  
\* 70084 79364: contig of 9281 bp in length  
\* 79365 79464: gap of 100 bp  
\* 79465 87765: contig of 8301 bp in length  
\* 87766 87865: gap of 100 bp  
\* 87866 97785: contig of 9920 bp in length  
\* 97786 97886: gap of 100 bp  
\* 97887 130019: contig of 32134 bp in length  
\* 130020 130119: gap of 100 bp  
\* 130120 144101: contig of 13982 bp in length  
\* 144102 144201: gap of 100 bp  
\* 144202 158508: contig of 14307 bp in length  
\* 158509 158608: gap of 100 bp  
\* 158609 174086: contig of 15478 bp in length  
\* 174087 174186: gap of 100 bp  
\* 174187 187332: contig of 13146 bp in length.  
\* Location/Qualifiers  
1. 187332  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
1. 23755  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
23856..24529  
/note="assembly\_fragment"  
24630..25795  
/note="assembly\_fragment"  
25996..26974  
/note="assembly\_fragment"  
27075..28604  
/note="assembly\_fragment"  
28705..30067  
/note="assembly\_fragment"  
30168..31088  
/note="assembly\_fragment"  
31189..32958  
/note="assembly\_fragment"  
33059..35230  
/note="assembly\_fragment"  
35331..37407  
/note="assembly\_fragment"  
37508..40202  
/note="assembly\_fragment"  
40303..43833  
/note="assembly\_fragment"  
43934..46356  
/note="assembly\_fragment"  
46457..48641  
/note="assembly\_fragment"  
48742..51680  
/note="assembly\_fragment"  
51781..54991  
/note="assembly\_fragment"  
55092..58578  
/note="assembly\_fragment"  
58679..62184  
/note="assembly\_fragment"  
62285..69983  
/note="assembly\_fragment"  
70084..79364  
/note="assembly\_fragment"  
79465..87765  
/note="assembly\_fragment"  
87866..97785  
/note="assembly\_fragment"

```

Pred. No.:      6 32e-106      Length:      187332
Score:          109.00      Matches:      109
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      60.89%      Indels: 0
DB:              2      Gaps: 0

US-09-701-674A-23 (1-179) x AC101292 (1-187332)

QY 42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArgLys 61
Db 82639 TCCAACTGCGAGAACGGGTCTCCACAGAAGGGTGGCGGTGGCTTAGCGAAGAGGAGGAG 82580

QY 62 AlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81
Db 82579 GCGCCCACTAAGAAAGCCCGCTCAGCGGGGTGTCAGCAGAGGGGACGAGTCCAGGCG 82520

QY 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101
Db 82519 AACGCGGCCCAATGCTGTCAGCGGGCCCGCATCGCGGTGCTGAGCAAGGCCCTTCTCCAGG 82460

QY 102 LeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThrLeu 121
Db 82459 CTCAGACCAACCTGCGCTGGGTGGTGGCCCGGACACCAAGCTCTCCAGTGGACACTTCT 82400

QY 122 ArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGlu 141
Db 82399 AGCGTGGCGTCCAGTACATCGCTACTTAAGCAGATCTCGGCCAACGACAGTACGAG 82340

QY 142 AsnGlyTyrlleHisProValAsnLeu 150
Db 82339 AACGGTTACATTCACCACTCAACCTG 82313

RESULT 14
LOCUS AC115183/c 241544 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-58J14, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC115183
VERSION AC115183.4 GI:25138798
KEYWORDS HTG; HTGS PRS1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mammus.
REFERENCE 1 (bases 1 to 241544)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,
Herrandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulsegged, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartre, M., Mahmoud, M., Malloy, K., Mangum, A.,

```

## TITLE

JOURNAL

## REFERENCE

AUTHORS

## TITLE

JOURNAL

## REFERENCE

AUTHORS

## TITLE

JOURNAL

## COMMENT

Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:22856693. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GSCC

Center clone name: CH230-58J14

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 214521 bases at least Q40

Consensus quality: 217222 bases at least Q30

Consensus quality: 219223 bases at least Q20

Estimated insert size: 218736; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 20:19:00 ; Search time 3794 Seconds

(without alignments)  
2231.117 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MGTGSLSDVEDLQEVEMLEC.....PESDLKEVVTASRLCGTTAS 179

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DRV=xlh  
-Q/cgna2\_1/USPTO.spool\_h/US09701674/runat\_15122004\_100440\_27505/app\_query.fasta\_1.327  
-DB=GenEmbl -QFMT=fastbp -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 @CGN 1 1 3731 @runat\_15122004\_100440\_27505 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	917	100.0	537	9	CR450293 Homo sapi
2	917	100.0	1272	6	BD211771 Proteins
3	917	100.0	3231	9	BC025697 Homo sapi
4	914	99.7	1254	9	AF035718 Homo sapi

5	913	99.6	1257	9	AF047419 Homo sapi
6	904	98.6	1259	6	CQ718038 Sequence
7	887	96.7	1202	10	AF047418 Mus muscu
8	887	96.7	1217	10	AF029753 Mus muscu
9	887	96.7	1220	10	AB009453 Mus muscu
10	887	96.7	1240	10	AF035717 Mus muscu
11	887	96.7	1246	10	BC053525 Mus muscu
12	887	96.7	1267	10	AF036945 Mus muscu
13	839	91.5	1164	5	BC073597 Xenopus ga
14	804	87.7	965	5	BC073597 Xenopus ga
15	804	87.7	1059	5	AY660871 Homo sapi
16	766	83.5	25186	9	AL356109 Human DNA
17	748.5	81.6	187332	2	AC101292 Mus muscu
18	746	81.4	241544	2	AC115183 Rattus no
19	646	70.4	513	10	AF061752 Rattus no
20	568	61.9	124070	5	EX530074 Zebrafish
21	568	61.9	231207	2	CR293516 Danio rer
22	561	61.2	189465	2	CR318665 Danio rer
23	514	56.1	621	9	BT007169 Homo sapi
24	514	56.1	621	12	BT008287 Synthetic
25	514	56.1	1917	6	CQ717187 Sequence
26	514	56.1	2031	9	BC006313 Homo sapi
27	514	56.1	2040	9	BC067827 Homo sapi
28	507	55.3	2382	6	AX281674 Sequence
29	495.5	54.0	606	10	AF108216 Mus muscu
30	495.5	54.0	1718	10	AF087035 Mus muscu
31	493.5	53.8	1716	9	AF087036 Homo sapi
32	482	52.6	1893	9	AF060154 Homo sapi
33	398	43.4	170668	9	AC104012 Homo sapi
34	398	43.4	198331	9	AC022867 Homo sapi
35	388.5	42.4	163698	10	AC115876 Mus muscu
36	388.5	42.4	181157	10	AC121789 Mus muscu
37	388.5	42.4	225399	2	AC113255 Rattus no
38	362.5	39.5	222122	2	CR450806 Danio rer
39	353	38.5	240	6	AR270465 Sequence
40	318	34.7	697	6	AX333347 Sequence
41	318	34.7	697	6	AX333763 Sequence
42	266	29.0	144007	2	AC141743 Apis mell
43	212.5	23.2	1066	3	AF000729 Drosophila
44	211.5	23.1	426	6	CQ722685 Sequence
45	211.5	23.1	1066	6	CQ582567 Sequence

## ALIGNMENTS

CR450293 537 bp mRNA linear PRI 18-MAY-2004  
Homo sapiens full open reading frame cDNA clone RZPD0834B11D for  
gene TCF21, transcription factor 21; complete cds; without  
stopcodon.

ACCESSION CR450293.1 GI:47496532  
VERSION CR450293  
KEYWORDS Full ORF shuttle clone, Gateway(TM), complete cds.  
SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 537)  
Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.  
Cloning of human full open reading frames in Gateway(TM) system  
entry vector (pDONR201)

## JOURNAL

2 (bases 1 to 537)  
Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.  
Direct Submission  
Submitted (18-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer  
Genomforschung GmbH, Im Neuenheimer Feld 580, D-69120 Heidelberg,  
Germany

## COMMENT

RZPD: RZPD0834B11D, ORFNO 76  
www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=RZPD0834B11D RZPDLIB:  
Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No.  
834

www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834  
www.rzpd.de/products/orclones/  
Contact: Ina Rolfs  
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,  
Heubnerweg 6, D-14959 Berlin, Germany  
Tel: +49 30 32639 101  
Fax: +49 30 32639 111  
www.rzpd.de

This clone is available from RZPD; contact RZPD (customer.service@rzpd.de) for further information. This CDS clone is a part of a collection of human full length expression clones generated by RZPD. This CDS has been cloned without stopcodon. This CDS has been inserted into pDONR201 via a BP Clonase(TM) reaction. Additional sequence has been added in front of the start codon (ATG): att. AAAAAA GGT GGC ACC CCT GGT CCA GGT (ATG). After the last codon additional sequence has been added: CCA GGC CCA CCA GGC GGC G in front of the 3' att site (AC CCA GCT TTC TM). Compared to the reference sequence NM\_003206 we did not find any amino acid exchanges. Clone distribution: <http://www.rzpd.de/products/orfclones/>.

FEATURES	SOURCE
----------	--------

```

1. 537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="RZPD034B11D"
/cdb_lib="Human Full ORF Clones Gateway (TM) - RZPD"
/lab_host="DH10B"
/db_host="Vector: pDONR201, Site_1: attPI, Site_2: attB2"
1. 537
/gene="TCF21"
1. 537
/gene="TCF21"
/codon_start=1
/protein_id="CAG29289.1"
/db_xref="GI:4796333"
/translation="MSTGSLSDVEDLQVEMLDGLKMDNSKEFYTSNTESE
EYQKRGGLGRKRPKPSISGVQSGEKGVORNAARERARMEVLKAL
TLTPWPPDPTLKLSDTLKASSYIAHLRLQLANDKYNGYHPVNLTPPFW
SDLKEVYTSARLGGTTAS"

```

## ORIGIN

Alignment Scores:		
Pred. No.:	1.65e-76	Length:
Score:	917.00	Matches:
Percent Similarity:	100.00%	Conservative:
Percent Local Similarity:	100.00%	Mismatches:
Query Match:	100.00%	Indels:
DB:	9	Gaps:

US-09-701-674A-23 (1-179) x CR450293 (1-537)

1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnValGluMetLeuGluCys	20
2		
3		
4	1	ATGTCCACCGCGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAAATGT
5		60
6		
7	21	AspGlyLeuLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu
8		40
9		
10	61	GACGGTTGAATCGATTTCGACNAGGAATTTGTGACTTCCNACGAGAGACCCGAGGAG
11		120
12	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg
13		60
14	121	AGCTCCAACTCGCAGATGGTCTCCCGAAGGGCGGGGGCGCTGGCGCAAGAGGAGG
15		180
16	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln
17		80
18	181	AAGCGGCCCAACCAAGACGCCCTCTAGCGGGGTGAGCAGAGGGGGAAGCAGGTCCAG
19		240
20	81	ArgAsnAlaAlaAlaAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer
21		100
22	241	CGAACCGCCCAACCGCGAGAGCGGGCCCGCATGCGATGCTGAGCAAGGCCCTTCTCC
23		300

101	ArgLeuIysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr	120
QY		
301	AGACTAAGACCAACCCCTGGCTGGGTGGCCCCCGACCAAGCTCTCCAAGCTGGACACG	360
Db		
121	LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr	140
QY		
361	CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGCGAGATCCTGGCTAAGCAGAAATAC	420
Db		
141	GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro	160
QY		
421	GAGAACGGGTACATTCACCGGTCACTGACGTGGCCCTTTATGGTGGCCGGGAACCC	480
Db		
161	GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer	179
QY		
481	GAGAGTCACTGAAAGAAAGTGGTGCACCGGACCGCCTTATGGGAACCAACCGCGTCC	537
Db		

RESULT 2  
BD211771

BD211771	LOCUS	BD211771	1272 bp	DNA	linear	PAT 17-JUL-2003
DEFINITION Proteins regulating gene expression.						

DEFINITION  
ACCESSION BD211771  
VERSION BD211771.1 GT:33021541  
KEYWORDS JP 2002517246-A/23.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

ORGANISM: *Homo sapiens*  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 1272)  
Mammalia; Eutheria; Primates; Catarrhini; Homidae.  
AUTHORS  
Lal, P., Yue, H., Tang, Y. T., Hillman, J. L., Bandman, O., Corley, N. C.,  
Guegler, K. J., Gorgone, G. A., Baughn, M. R., Patterson, C. and Lu, D. A. M.  
TITLE  
Proteins regulating gene expression  
Patent: JP 2002517245-A 23 18-JUN-2002  
JOURNAL

COMMENT	PN	JP	18-JUN-2002	PO	18-JUN-2002
OS Homo sapiens (human)	PN	JP 2002517246-A/23			
INCYTE PHARMACEUTICALS INC					
FACTORY OF 2002517246-A/23					18-JUN-2002

PD 18-JUN-2002  
PF 11-JUN-1999 JP 200053586  
PR 12-JUN-1998 US 60/089029, 29-JUL-1998 US 60/094575 PR  
14-OCT-1998 US 60/104624  
PI PRSETI LAL, HENRY YUE, Y TOM TANG, JENNIFER L HILLMAN, OLGA PI  
BANDMAN.

## FEATURES

```

1. .1272
source
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      4,86e-76      Length:      1272
Score:          917.00        Matches:      179
Percent Similarity: 100.0%    Conservative: 0
Local Similarity:  100.0%    Mismatches:  0
Query Match:       100.0%    Indels:       0
DB:                6        Gaps:         0

```

US-09-701-674A-23 (1-179) x BD211771 (1-1272)

1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnValGluMetLeuGluCys 20  
268 ATGTCCACCGGCTCCCTCCGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTTGGAATGT 327  
21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40



```

Db 380 AGTCCAACTGCAGAGTGGTCTCCCAAGAGGCGCGCGCTGGCAAGAGGAGG 439
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyValGln 80
Db 440 AAGGCGCCCAACCAAGAGAGCCCTTGAGCGGGTTCAGCAGAGGAGGAGGTCAG 499
QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 500 CGCAACGCGCCCAACGCGAGAGCGGCGCGCATCGAGAGTGTGAGCAAGGCTTCTCC 559
QY 101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysSerLysLeuAspThr 120
Db 560 AGACTCAAGACCACTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 619
QY 121 LeuArgLeuAlaSerGlyTrpIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db 620 CTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGCAGATCTGCTTAACGACAAATAC 679
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db 680 GAGACGCGGTACATTCACCGCGTCAACCTGAGCGTGGCGCTTTATGGTGGCGGGAACCC 739
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 740 GAGAGTGACCTGAAGAGTGGTGACCGCGCGCTTATGTGGAACCAACCGCGTCC 796

RESULT 4
AF035718 1254 bp mRNA linear PRI 05-OCT-1998
LOCUS Homo sapiens mesoderm-specific basic-helix-loop-helix protein
DEFINITION (POD1) mRNA, complete cds.
ACCESSION AF035718
VERSION 98175875
KEYWORDS /db_xref="taxon:9606"
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1254)
AUTHORS Quaggin,S.E., Vanden Heuvel,G.B. and Igarashi,P.
TITLE Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed in mesenchymal and glomerular epithelial cells in the developing kidney
JOURNAL Mech. Dev. 71 (1-2), 37-48 (1998)
MEDLINE 98175875
PUBMED 9507058
REFERENCE 2 (bases 1 to 1254)
AUTHORS Quaggin,S.E., Vanden Heuvel,G.B. and Igarashi,P.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1997) Internal Medicine, Yale University, 333 Cedar Street, New Haven, CT 06520-8029, USA

FEATURES
source 1..1254
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /chromosomes="6"
gene 1..1254
    /gene="POD1"
CDS 261..800
    /gene="POD1"
    /note="Pod-1"
    /product="mesoderm-specific basic-helix-loop-helix protein"
    /protein_id="AAC62514.1"
    /db_xref="GI:2745887"
    /translation="WSTGSLSDVLEQVEMLECDGLKMDSNKEFVTSNESTERSNC
    ENGSPQKGGGKRRAPTKVPSGVSQSGKQVORNAANARARARVLSKAFSRL
    KTLFWPPDTKLKSLDTRLASSYIAHLRLANDKYNGIYHFNLTWPFVWAGKP
    ESDLKEVNTASLRCGTAS"
misc_feature 492..653

```

```

/ gene="POD1"
/ note="encodes basic-helix-loop-helix domain"

ORIGIN
Alignment Scores: 9.08e-76 Length: 1254
Pred. No.: 914.00 Matches: 178
Score: 914.00 Conservative: 1
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 99.44% Indels: 0
Query Match: 99.67% Caps: 0
DB: 9

US-09-701-674a-23 (1-179) x AF035718 (1-1254)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 261 ATGTCCACCGGCTCCCTCAGCGATGTGGAGCAGCTTCAAGAGGTGGAGATGTTGGAATGT 320
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 321 GACGGGTGAAATGGATTGAAACAGGAATTTGTGACTTCCAAACGAGAGCACCGAGGAG 380
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db 381 AGCTCCAACTGCAGAGATGGTCTCTCCCAAGAGGCGCGCGCTGGGCAAGAGGAGG 440
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 441 AGGCGCGCCCAACCAAGAGAGCGCCCTTGAGCGGGTTCAGCAGAGGAGGAGGTCAG 500
QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 501 CGCAACGCGCCCAACGCGAGAGCGGCGCGCATCGAGAGTGTGAGCAAGGCTTCTCC 560
QY 101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysSerLysLeuAspThr 120
Db 561 AGACTCAAGACCACTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 620
QY 121 LeuArgLeuAlaSerGlyTrpIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db 621 CTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGCAGATCTGCTTAACGACAAATAC 680
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db 681 GAGAACGCGGTACATTCACCGCGTCAACCTGAGCGCTTTATGGTGGCGGGAACCC 740
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 741 GAGAGTGACCTGAAGAGTGGTGACCGCGCGCTTATGTGGAACCAACCGCGTCC 797

RESULT 5
AF047419 1257 bp mRNA linear PRI 06-OCT-1998
LOCUS Homo sapiens epicardin mRNA, complete cds.
DEFINITION AF047419
ACCESSION AF047419
VERSION AF047419.1 GI:2921852
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1257)
AUTHORS Robb,L.L., Mifsud,L., Hartley,L., Biben,C., Copeland,N.G.,
Gilbert,D.J., Jenkins,N.A. and Harvey,R.P.
TITLE epicardin: A novel basic helix-loop-helix transcription factor gene expressed in epicardium, branchial arch myoblasts, and mesenchyme of developing lung, gut, kidney, and gonads
JOURNAL Dev. Dyn. 213 (1), 105-113 (1998)
MEDLINE 98401941
PUBMED 9733105
REFERENCE 2 (bases 1 to 1257)
AUTHORS Robb,L.L. and Mifsud,L.
TITLE Direct Submission

```

Sept, bad date

JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Melbourne, VIC 3050, Australia

FEATURES Location/Qualifiers

source 1..1257  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

CDS 260..799  
 /function="basic helix-loop-helix transcription factor"  
 /note="expressed in embryonic epicardium and mesenchyme-derived tissues of lung, gut, kidney, and gonad"

/codon\_start=1  
 /product="epicardin"  
 /protein\_id="AAC62532.1"  
 /db\_xref="GI:2921853"  
 /translation="MSTGSLSDVEDLQVEMLECDGLKMDSNKPFVTSNESTEESSNC  
 ENGSPKGGGLGKPKAPTKPSPLSGVSOEGKQVORNTANAREARVLSKAPSRLL  
 KTLPPWPDPTKLSKLDLRLASSYIAHRLQILANDKYENGIHPNLTWFPWVAGKP  
 ESDLKEVVTASRLCGTTAS"

## ORIGIN

Alignment Scores:  
 Pred. No.: 113e-75 Length: 1257  
 Score: 913.00 Matches: 178  
 Percent Similarity: 99.44% Conservative: 0  
 Best Local Similarity: 99.44% Mismatches: 1  
 Query Match: 99.56% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x AF047419 (1-1257)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 260 ATGTCACCGGCTCCCTCAGCATGTGGAGACCTTCAAGAGTGGAGATGTTGAATGT 319  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 DB 320 GACGGGTGAAATGGAATTCACACAGGAATTTGTGACTTCCACGAGAGCCCGGAG 379  
 QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 DB 380 AGTCCCACTGCCAGAAATGGGTCTCCACAGAGGGCGGGCGGCTGGGCAAGAGGAG 439  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 DB 440 AAGCGCCCAACAGAGAGCCCTCAGCGGGGTGAGCGAGGGGAGGAGGTCAG 499  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 DB 500 CGCAACACCCCAAGCGCGAGAGCGGCCCGCATGCGAGTGTGAGCAAGGCTTCTCC 559  
 QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 DB 560 AGACTCAAGACCACTCCCTGGGTGCCCGCCGACCAAGCTCTCCAAGCTGGACACG 619  
 QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 DB 620 CTGAGCTGGCTCCAGCTACATCGCCACTTGAGCGAGATCTGCTGACGACCAATAC 679  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 DB 680 GAGAACGGGTACATTACCCGGTCAACTGACGTGGCCCTTTATGTGGCCGGGAACCC 739  
 QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 DB 740 GAGAGTGACCTGAAGAAGTGGTGACCGCGAGCCGCTTATGTGAACACCCCGCTCC 796

## RESULT 6

CQ718038 1259 bp DNA linear PAT 03-FEB-2004  
 LOCUS Sequence 3972 from Patent WO02068579.  
 DEFINITION

ACCESSION CQ718038  
 VERSION CQ718038.1  
 KEYWORDS GI:42278895

## SOURCE

ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
 TITLE Kits, such as nucleic acid arrays, comprising a majority of  
 humanexons or transcripts, for detecting expression and other uses  
 thereof

## JOURNAL

Patent: WO 02068579-A 3972 06-SEP-2002;

## FEATURES

source FE Corporation (NY) (US)  
 Location/Qualifiers  
 1..1259  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
 Pred. No.: 7.81e-75 Length: 1259  
 Score: 904.00 Matches: 179  
 Percent Similarity: 99.44% Conservative: 0  
 Best Local Similarity: 99.44% Mismatches: 0  
 Query Match: 98.58% Indels: 1  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CQ718038 (1-1259)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 264 ATGTCACCGGCTCCCTCAGCATGTGGAGACCTTCAAGAGTGGAGATGTTGAATGT 323  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 DB 324 GACGGGTGAAATGGAATTCGACACAGGAATTTGTGACTTCCACGAGAGCAGCGAGG 383  
 QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 DB 384 AGTCCCACTGCCAGAAATGGGTCTCCACAGAGGGCGGGCGGCTGGGCAAGAGGAG 443  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 DB 444 AAGCGCCCAACAGAGAGCCCTCAGCGGGGTGAGCGAGGGGAGGAGGTCAG 503  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSe 100  
 DB 504 CGCAACCGCCCAACCGCGAGAGCGGGCCCGCATGCGAGTGTGAGCAAGGCTTCTC 563  
 QY 100 rArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspTh 120  
 DB 564 CAGACTCAAGACCACTCCCTGGGTGCCCGCCGACCAAGCTCTCCAAGCTGGACAC 623  
 QY 120 rLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTy 140  
 DB 624 GCTCAGGTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTGCTGACGACCAATA 683  
 QY 140 rGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPr 160  
 DB 684 CGAGAACGGGTACATTACCCGGTCACTGACGTGGCCCTTTATGTGGCCGGGAAC 743  
 QY 160 oGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 DB 744 CGAGAGTGACCTGAAGAAGTGGTGACCGCGAGCCGCTTATGTGAACACCCCGGTCC 801

## RESULT 7

AF047418 1202 bp mRNA linear ROD 06-OCT-1998  
 LOCUS Mus musculus epicardin mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF047418  
 VERSION AF047418.1  
 GI:2921850

**KEYWORDS**  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**AUTHORS** 1 (bases 1 to 1202)  
**TITLE** Robb, L., Mifsud, L., Hartley, L., Biben, C., Copeland, N.G., Gilbert, D.J., Jenkins, N.A. and Harvey, R.P.  
**epicardin: A novel basic helix-loop-helix transcription factor gene expressed in epicardium, branchial arch myoblasts, and mesenchyme of developing lung, gut, kidney, and gonads**  
**JOURNAL** Dev. Dyn. 213 (1), 105-113 (1998)  
**MEDLINE** 98401941  
**PUBMED** 9733105  
**REFERENCE** 2 (bases 1 to 1202)  
**AUTHORS** Robb, L. and Mifsud, L.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Melbourne, VIC 3050, Australia  
**FEATURES** Location/Qualifiers  
**source**  
1. .1202  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Balb/c"  
/db\_xref="taxon:10090"  
199. .738  
/function="basic helix-loop-helix transcription factor"  
/notes="expressed in embryonic epicardium and mesenchyme-derived tissues of lung, gut, kidney, and gonad"  
/codon\_start=1  
/product="epicardin"  
/protein\_id="AAC6253.1"  
/db\_xref="GI:2921851"  
/translation="MSTGLSDVEDLQVEMLDCSLKVDNKEFGTSGNESTEGSNC  
ENGSPQKGGGLGKRRKAPTKKSPGVSQEGQVORNAARERARVLSKAFSRL  
KTLPLWPPDTKLKSLDTLRASSYIAHLRQILANDKYENGIHPVNLTPFWVAGPK  
ENDLKEVTVANRLCGTTAS"  
**CDS**  
1. .1202  
/codon\_start=1  
/product="epicardin"  
/protein\_id="AAC6253.1"  
/db\_xref="GI:2921851"  
/translation="MSTGLSDVEDLQVEMLDCSLKVDNKEFGTSGNESTEGSNC  
ENGSPQKGGGLGKRRKAPTKKSPGVSQEGQVORNAARERARVLSKAFSRL  
KTLPLWPPDTKLKSLDTLRASSYIAHLRQILANDKYENGIHPVNLTPFWVAGPK  
ENDLKEVTVANRLCGTTAS"  
**ORIGIN**  
Alignment Scores:  
Pred. No.: 2,83e-73 Length: 1202  
Score: 887.00 Matches: 172  
Percent Similarity: 98.32% Conservative: 4  
Best Local Similarity: 96.09% Mismatches: 3  
Query Match: 96.73% Indels: 0  
DB: 10 Gaps: 0  
US-09-701-674A-23 (1-179) x AF047418 (1-1202)  
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
DB 199 AUGTCCATCGCTCCCTCAGCATGATGAGAACCTTCAAGAGTGGAGATCGTCACTGT 259  
QY 21 AspGlyLeuLeuMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
DB 259 GACTCCCTGAAGTGGACTCCAACAGGAGTTTGGAACTTCCACGAGACACCGAGGAG 318  
QY 41 SerSerAsnGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
DB 319 GGCTTCAACTGGAGAACCGGTCTCCACAGAGGGTCCGGTGGCTAGGCAAGAGGAGG 378  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
DB 379 AAGGCCCCCACTAAGAAACCCCGCTCAGCGGGGTACCCAGGAGGGCAGAGTCCAG 438  
QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
DB 439 CGCAACGGCCCAATGCTCTGTCGAGCGGGCCCGCATGCGGGTCTGAGCAAGGCCTTCTCC 498  
QY 101 ArgGluLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120

*Spe; bad date*

*1026a*

**Db** 499 AGGCTCAAGACCAACCTGCTGGTGCCTCCCGCACCAAGCTCTCCAAAGCTGGACACT 558  
**Qy** 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
**Db** 559 CTCAGGCTGGCTCAGCTACATCGCTTAAAGCAGATCTTGGCCAAACGAAAGTAC 618  
**Qy** 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
**Db** 619 GAGAAAGGTTACATTACCCAGTCAACCTGACGTGGCCCTTTATGTGGTGGCGGCAACCA 678  
**Qy** 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
**Db** 679 GAGAACTGACCTGAAGGAAGTGGTACCGCCCAACCGCTTGTGTGGAACCTACAGCATCC 735  
**RESULT 8**  
**AF029753** 1217 bp mRNA linear ROD 25-FEB-2000  
**LOCUS** Mus musculus basic helix-loop-helix factor Cor1 mRNA, complete cds.  
**DEFINITION** AF029753  
**ACCESSION** AF029753.1 GI:2613018  
**VERSION** AF029753.1  
**KEYWORDS** Mus musculus (house mouse)  
**SOURCE** Mus musculus  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**AUTHORS** 1 (bases 1 to 1217)  
Hidai, H., Bardales, R., Goodwin, R., Quettermous, T. and Quettermous, E.E.  
**TITLE** Cloning of capulin, a basic helix-loop-helix factor expressed in progenitor cells of the pericardium and the coronary arteries  
**JOURNAL** Mech. Dev. 73 (1), 33-43 (1998)  
**MEDLINE** 98213629  
**PUBMED** 9545526  
**REFERENCE** 2 (bases 1 to 1217)  
**AUTHORS** Quettermous, T., Quettermous, E.E., Hidai, H., Padel, B., Boutet, S.C. and Bardales, R.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (10-OCT-1997) Department of Medicine, Division of Cardiology, Stanford University, Falk Bldg, 300 Pasteur Drive, Stanford, CA 94305, USA  
**FEATURES** Location/Qualifiers  
1. .1217  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="pericardium and coronary arteries"  
1. .1217  
/genes="Cor1"  
/genes="Cor1"  
/function="transcription factor"  
/codon\_start=1  
/product="basic helix-loop-helix factor Cor1"  
/protein\_id="AAB84256.1"  
/db\_xref="GI:2613019"  
/translation="MSTGLSDVEDLQVEMLDCSLKVDNKEFGTSGNESTEGSNC  
ENGSPQKGGGLGKRRKAPTKKSPGVSQEGQVORNAARERARVLSKAFSRL  
KTLPLWPPDTKLKSLDTLRASSYIAHLRQILANDKYENGIHPVNLTPFWVAGPK  
ENDLKEVTVANRLCGTTAS"  
**gene**  
**CDS**  
1. .1217  
/codon\_start=1  
/product="basic helix-loop-helix factor Cor1"  
/protein\_id="AAB84256.1"  
/db\_xref="GI:2613019"  
/translation="MSTGLSDVEDLQVEMLDCSLKVDNKEFGTSGNESTEGSNC  
ENGSPQKGGGLGKRRKAPTKKSPGVSQEGQVORNAARERARVLSKAFSRL  
KTLPLWPPDTKLKSLDTLRASSYIAHLRQILANDKYENGIHPVNLTPFWVAGPK  
ENDLKEVTVANRLCGTTAS"  
**ORIGIN**  
Alignment Scores:  
Pred. No.: 2,88e-73 Length: 1217  
Score: 887.00 Matches: 172  
Percent Similarity: 98.32% Conservative: 4  
Best Local Similarity: 96.09% Mismatches: 3  
Query Match: 96.73% Indels: 0  
DB: 10 Gaps: 0  
US-09-701-674A-23 (1-179) x AF029753 (1-1217)  
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20





[illegible]

gene /clone\_lib="NCI\_CGAP\_Co24"  
 /lab\_host="DH10B"  
 /notes="vector: pCMV-SPORT6"  
 1. .1246  
 /genes="Tcf21"  
 /notes="synonyms: epc, Pod1; capsulin, epicardin, Pod-1"  
 /db\_xref="LocusID:21412"  
 /db\_xref="MGI:1202715"  
 233..772  
 /gene="Tcf21"  
 /codon\_start=1  
 /product="transcription factor 21"  
 /protein\_id="AAH53525.1"  
 /db\_xref="GI:31566110"  
 /db\_xref="LocusID:21412"  
 /db\_xref="MGI:1202715"  
 /translation="MSTGSLSDVEDLOEVMLDCDSLKVDNKEFGTSNESTEGSNC  
 ENGSPQGRGLGKRKAPTKKSPGSGVQKQVQRNAANARERAMRVLSKAFSRL  
 KTLTPWPVPTKLSKLDTLRLASSYTAHLRQLANDKYENGYIHPVNLTPFWVAGRP  
 ENDLKEVVTANRLCGTTAS"

## ORIGIN

Alignment Scores:  
 Pred. No.: 2,97e-73 Length: 1246  
 Score: 887.00 Matches: 172  
 Percent Similarity: 98.32% Conservative: 4  
 Best Local Similarity: 96.09% Mismatches: 3  
 Query Match: 96.73% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x BC053525 (1-1246)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 233 ATGTCCACTGGCTCCCTCAGCGATGTAGAAGACCTTCAAGAGGTGGAGATGCTGGACTGT 292  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 293 GACTCCCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 352  
 QY 41 SerSerAsnGlySerProGlnLysGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60  
 Db 353 GGCTCCACTGCGAGAACGGGTCTCCACAGAGGGTCCGGTGGTGGTGGTGGTGGTGGTGGTGGT 412  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 413 AAGCGCCCACTAAGAAAGCCCGCTCAGCGGGGTCCAGCGAGGGTCAAGCGAGGTCCAG 472  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 473 CGCAACGGGCCCAATGCTGTGAGCGGGCCCGCATCGGGTGTGTGAGCAAGCCCTTCTCC 532  
 QY 101 ArgLeuLysThrThrLeuProTTPValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 533 AGGCTCAAGACCACTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 592  
 QY 121 LeuArgLeuAlaSerSerThrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 593 CTGAGGT 652  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTTPProPheMetValAlaGlyLysPro 160  
 Db 653 GAGAACGGTTACATTCAACCCAGTCAACCTGACGTGGCCCTTTATGGTGGCCGCAACCA 712  
 QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 713 GAGATGACCTGAAGAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 769

## RESULT 12

AF036945 1267 bp mRNA linear ROD 15-JUN-1998  
 LOCUS Mus musculus capsulin mRNA, complete cds.  
 DEFINITION AF036945  
 ACCESSION

AF036945.1 GI:2695694  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1267)  
 Lu, J., Richardson, J.A. and Olson, E.N.  
 Capsulin: a novel BHLH transcription factor expressed in epicardial  
 progenitors and mesenchyme of visceral organs  
 Mech. Dev. 73 (1), 23-32 (1998)  
 9545521  
 98213628  
 2 (bases 1 to 1267)  
 Lu, J. and Olson, E.N.  
 Direct Submission  
 Submitted (04-DEC-1997) Molecular Biology and Oncology, UT  
 Southwestern Medical Center, 6000 Harry Hines Blvd., NAB.510,  
 Dallas, TX 75235-9148, USA  
 Location/Qualifiers  
 1..1267  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 264..803  
 /function="BHLH transcription factor"  
 /codon\_start=1  
 /product="capsulin"  
 /protein\_id="AAC23537.1"  
 /db\_xref="GI:2695694"  
 /translation="MSTGSLSDVEDLOEVMLDCDSLKVDNKEFGTSNESTEGSNC  
 ENGSPQGRGLGKRKAPTKKSPGSGVQKQVQRNAANARERAMRVLSKAFSRL  
 KTLTPWPVPTKLSKLDTLRLASSYTAHLRQLANDKYENGYIHPVNLTPFWVAGRP  
 ENDLKEVVTANRLCGTTAS"

Alignment Scores:  
 Pred. No.: 3.03e-73 Length: 1267  
 Score: 887.00 Matches: 172  
 Percent Similarity: 98.32% Conservative: 4  
 Best Local Similarity: 96.09% Mismatches: 3  
 Query Match: 96.73% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x AF036945 (1-1267)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 264 ATGTCCACTGGCTCCCTCAGCGATGTAGAAGACCTTCAAGAGGTGGAGATGCTGGACTGT 323  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 324 GACTCCCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 383  
 QY 41 SerSerAsnGlySerProGlnLysGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60  
 Db 384 GGCTCCCACTCGGAGAACGGGTCTCCACAGAGGGTCCGGTGGCTTAGCAAGAGGAGG 443  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 444 AAGCGCCCACTAAGAAAGCCCGCTCAGCGGGGTCCAGCAGGAGGGGCAAGAGGTCCAG 503  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 504 CGCAACGGGCCCAATGCTGTGAGCGGGCCCGCATCGGGTGTGTGAGCAAGCCCTTCTCC 563  
 QY 101 ArgLeuLysThrThrLeuProTTPValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 564 AGGCTCAAGACCACTCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 623  
 QY 121 LeuArgLeuAlaSerSerThrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 624 CTCAGGTGGGTCCAGCTACATCGCTACCTTAAGGCGAGATCCTGGCCCAACGACAGATAC 683

141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 684 GAGACGGTTACATTACCCAGTCAACCTGACGTGGCCCTTTATGTGGCCGGCAACCA 743

161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 744 GAGAAATGACCTGAAGGAAGTGGTGACCGCAACCGCTTGTGTGGAACACTACACATCC 800

RESULT 13  
 Bx932800  
 LOCUS Bx932800 1164 bp mRNA linear VRT 30-MAR-2004  
 DEFINITION Gallus gallus finished cDNA, clone CHEST265h24.  
 ACCESSION Bx932800  
 VERSION Bx932800.2 GI:46017571  
 KEYWORDS  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 1 (bases 1 to 1164)  
 Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E., Croning, M.D.R., Davies, R.M., Francis, M.D., Grafton, D.V., Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R., Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G., Tickie, C., and Wilson, S.A.  
 Direct Submission  
 Submitted (29-MAR-2004) Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: chickest@ms.unist.ac.uk  
 On Apr 1, 2004 this sequence version replaced gi:41633328.  
 BSRRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA sequencing project.  
 This sequence is from the BSRRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection, from a library constructed by Elizabeth Bosch. cDNA was prepared from RNA extracted from ovary, normalised, and poly A-trimmed. EcoRI-NotI cut cDNA was then ligated into the vector. Vector: pBluescript II KS(+); Site\_1: EcoRI; Site\_2: NotI Host: Escherichia coli DH10B.

FEATURES  
 source 1. 1164  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer"  
 /db\_xref="taxon:9031"  
 /clone="CHEST265h24"  
 /clone\_lib="CSQRBN19"  
 /dev\_stage="adult"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 8 13e-69 Length: 1164  
 Score: 839.00 Matches: 164  
 Percent Similarity: 93.85% Conservative: 4  
 Best Local Similarity: 91.62% Mismatches: 11  
 Query Match: 91.49% Indels: 0  
 DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x Bx932800 (1-1164)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 167 ATGTCACCTGGGTCCTCGATGATGGAAGATCTCGAAGAGTGGAAATGCTGAGTGC 226

21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 227 GATGGCTGAAATGATACATAACAGAGATTCGGGGCGCTCCAAAGAGACACAGAG 286

41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 287 GGATCCAAATGGCGAGAATGGCTCCCTCAGAAAGGAGAGGGCGCTCGGGAAGAGGAA 346

QY 61 LysAlaProThrIlyLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 347 AAGGCTCCCCCAAGAGAGCCCTTTAAATGGTGTGAGCCAGGAGAAAGCAGGTCAG 406

QY 81 ArgSerAlaAlaSerAlaArgGluArgAlaSerMetArgValLeuSerLysAlaPheSer 100  
 407 AGAAACCTGCCAACGCAAGGAGGAGGCAAGATAGGGTCTTTAGCAAGCTTCTCC 466

QY 101 ArgLeuLysThrThrLeuProTyrValProAspThrLysLeuSerLysLeuAspThr 120  
 467 AGGCTTAAAGACCACTCCCTGGTGGCCGCCAGACACCAAGCTTCCAACTGGACACC 526

QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaSerAspLysTyr 140  
 527 TTGAGGCTGGCTCCAGCTACATCGCACCTGAGGCATAATCTGCGCAACGCAAGTAC 586

QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 587 GAAATGGCTACATCCATCCAGTCAACCTGACCTGGCTTTATGTAGTCGCGCAACCC 646

QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 647 GAGAGTGACCTGAAGGAAGTGGTGAACACCAACCGCTGTGCGCGCCGCGGATCC 703

RESULT 14  
 Bx073597  
 LOCUS Bx073597 965 bp mRNA linear VRT 03-AUG-2004  
 DEFINITION Xenopus laevis transcription factor 21, mRNA (cDNA clone MGC:82899 IMAGE:5512805), complete cds.  
 ACCESSION Bx073597  
 VERSION Bx073597.1 GI:49256164  
 KEYWORDS  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Xenopus  
 1 (bases 1 to 965)  
 Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W. and Richardson, P.  
 Genetic and genomic tools for Xenopus research: The NIH Xenopus Initiative  
 Dev. Dyn. 225 (4), 384-391 (2002)  
 12454917  
 2 (bases 1 to 965)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Suetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smaluk, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 3 (bases 1 to 965)  
 Klein, S. and Gerhard, D.S.  
 Direct Submission  
 Submitted (14-JUN-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD

TITLE  
 JOURNAL  
 PUBMED  
 REFERENCE  
 AUTHORS

20892-7510, USA  
 REMARK NIH-MGC Project  
 COMMENT Contact: XGC help desk  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Dr. Igor Dawid  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 BC Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca  
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,  
 Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth  
 Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,  
 Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,  
 Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lilia Prabhu,  
 Parvaneh Saeeedi, JR Santos, Angelique Schnerch, Ursula Skalska,  
 Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacques  
 Schein, Asim Siddiqui, Rob Holt, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAK plate: 154 Row: n Column: 2  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Similarity but not  
 identity to protein.

## FEATURES

source

1. .965  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="MGC:82899 IMAGE:5512805"  
 /tissue\_type="Spleen, adult Xenopus"  
 /clone\_lib="NICHG\_XGC\_Spl"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 1. .965  
 /gene="Tcf21"  
 /note="synonym: MGC82899"  
 /db\_xref="LocusID:444386"  
 24. .563  
 /gene="Tcf21"  
 /codon\_start=1  
 /product="Tcf21 protein"  
 /protein\_id="AAH73597.1"  
 /db\_xref="GI:49256165"  
 /db\_xref="LocusID:444386"  
 /translation="MSTGSLSDVEDFQDMLECDGKILDPNKEFGISNDNESSTC  
 DNGSPKRGSTGKRKSKSKSPKPLGTINQEGKQVORNAARARVLSKAFSL  
 KITLPWPDPDKLSKIDTLRLASSYIAHLRQILANDKYENGYIHPVNLTPFWMVAGKP  
 ENDLKEVSTSLRCGPTAS"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.18e-65 Length: 965  
 Score: 804.00 Matches: 154  
 Percent Similarity: 92.74% Conservative: 12  
 Best Local Similarity: 86.03% Mismatches: 13  
 Query Match: 87.68% Indels: 0  
 DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BC073597 (1-965)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 24 ATGTCCACCGGTCTCTCAGTGATGGAGGACTTCCAGACATGGAAATGTTGGAATGT 83  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGlySerThrGluGlu 40  
 Db 84 GATGGCATCAAACTGATGCTCTAATAAGAAATTTGGCATATCAACACGACGAAATGAGGAG 143  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyValGlyGlyGlyLysArgArg 60  
 Db 144 AGTTCAACTGGACACACGGTTCCACCAAAAAGGTAGAGGACTTCGGGCGAAGAGGAGA 203

QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluLysGlnValGln 80  
 Db 204 AAAGCTTCAGCAAGAAAGTCCACTGGCGACCATCAACAGGAGGCGAGAGGTCAG 263  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 264 AGGAATGACCAATGCCAGGAGAGAGCCAGGATGAGGTGCTAGCAAAAGCCTTTCT 323  
 QY 101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysLeuSerLysLeuAspThr 120  
 Db 324 AGGTAAGACCATTTTGGCTTGGTGGCTCCAGACACCAAACTTCCAGAGCTGGACACT 383  
 QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 384 CTGGCGTTGCACTAGTACATAGCTACCTGCTGGTCAAAATCTAGCAATGACAAGTAT 443  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTyrProPheMetValAlaGlyLysPro 160  
 Db 444 GAAATAGCTACATCCATCCAGTTAAGTACCTTGGCCCTTTAAGTGGTGGCTGGAAACCA 503  
 QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 504 GAAACACCTCAAGGAAGTGGTGAGCACTAGCGGATTTGTGTGCCACCGCATCT 560  
 RESULT 15  
 AY660871 1059 bp mRNA linear VRT 20-JUL-2004  
 LOCUS Xenopus laevis transcription factor 21 (Tcf21) mRNA, complete cds.  
 DEFINITION  
 ACCESSION AY660871  
 VERSION AY660871.1 GI:50313157  
 KEYWORDS  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 Xenopodinae; Xenopus; Xenopus.  
 REFERENCE  
 1 (bases 1 to 1059)  
 Eid, S.R. and Brandli, A.W.  
 Essential function of the basic helix-loop-helix transcription  
 factor Tcf21 (capsulin/epicardin/Pod1) for Xenopus pronephric  
 kidney organogenesis  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 1059)  
 Eid, S.R. and Brandli, A.W.  
 Direct Submission  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (21-JUN-2004) Institute of Pharmaceutical Sciences, Swiss  
 Federal Institute of Technology (ETHZ), Winterthurerstrasse 190,  
 Zurich CH-8057, Switzerland  
 FEATURES  
 Location/Qualifiers  
 1. .1059  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="Pod1-12b"  
 /tissue\_type="kidney"  
 /dev\_stage="adult"  
 1. .1059  
 /gene="Tcf21"  
 134. .673  
 /gene="Tcf21"  
 /note="basic helix-loop-helix transcription factor;  
 capsulin; epicardin; podocyte-expressed 1; Pod1; Pod-1"  
 /codon\_start=1  
 /product="transcription factor 21"  
 /protein\_id="AA74527.1"  
 /db\_xref="GI:50313158"  
 /translation="MSTGSLSDVEDFQDMLECDGKILDPNKEFGISNDNESSTC  
 DNGSPKRGSTGKRKSKSKSPKPLGTINQEGKQVORNAARARVLSKAFSL  
 KITLPWPDPDKLSKIDTLRLASSYIAHLRQILANDKYENGYIHPVNLTPFWMVAGRP  
 ENDLKEVSTSLRCGPTAS"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,32e-65 Length: 1059  
 Score: 804.00 Matches: 154  
 Percent Similarity: 92.74% Conservative: 12  
 Best Local Similarity: 86.03% Mismatches: 13  
 Query Match: 87.68% Indels: 0  
 DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x AY660871 (1-1059)

```

Qy      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
      |||
Db      134 ATGTCACCGGTTCTCTAGTGTGGAGGACTTCCAAAGACATGGAATGTGGAATGT 193
      |||
Qy      21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
      |||
Db      194 GATGGCATCAAACTGGATCCTTAATAAGAAATTGGCATATCCACAGCAGCAATGAGGAG 253
      |||
Qy      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60
      |||
Db      254 AGTTCACCTGGCAGCAACCGTTGCGCCAAAAAAGGTAGAGGACTTCGGGCCAAGAGGAGA 313
      |||
Qy      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
      |||
Db      314 AAAGCCTCCAGCAAGAAAGTCCACTGGGCACCATCAACAGGAGGGCAAGCAGGTCCAG 373
      |||
Qy      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
      |||
Db      374 AGGAATGCAGCCAAATCCAGGAGAGAGCCAGGATGAGGGTGCTAAGCAAAAGCCTTTTCT 433
      |||
Qy      101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120
      |||
Db      434 AGCTAAAGACCACCTTTGCTTGGGTGGCTCCAGACCCAAACTTTCCAGCTGGACACT 493
      |||
Qy      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
      |||
Db      494 CTGGGCTTGCATCTAGTTACATAGCTACCTGCGTCAAAATCTAGCCAATGACAAGTAT 553
      |||
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
      |||
Db      554 GAAATGGCTACATCCATCCAGTTAACCTGACTTGGCCCTTTTATGGTGGCTGGAACCA 613
      |||
Qy      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
      |||
Db      614 GAAACGACCTCAAGGAAGTGGTGAGCACTAGCCGATTGTGTGTCCTCCACCGGCATCT 670
  
```

Search completed: December 16, 2004, 23:15:58

Job time : 3803 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 16, 2004, 15:39:22 ; Search time 682 Seconds  
(without alignments)  
7930.709 Million cell updates/sec

Title: US-09-701-674A-54  
Perfect score: 1272  
Sequence: 1 tctacggccacgactctggg.....attgctaaaaa 1272

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1272	100.0	1272	3	Aaz57861 Protein r
2	581	45.7	697	6	Abi65519 Lung can
3	581	45.7	697	6	Abi65935 Lung can
4	581	45.7	697	6	Abx64784 Human ben
5	525.6	41.3	632	6	Abc55427 Human ova
6	466.2	36.7	471	6	Abi66121 Lung can
7	466.2	36.7	471	6	Abi65571 Lung can
8	466.2	36.7	471	6	Abx64115 Human ben
9	434.2	34.1	493	9	Ach21857 Human adu
10	297.2	23.4	482	9	Ach21857 Human adu
11	246.8	19.4	271	3	Aac08260 Human sec
12	236.2	18.6	265	10	Adk61063 Ovarian c
13	236.2	18.6	265	12	Adm32307 Human can
14	213	16.7	240	10	Aca56430 Chicken s
15	213	16.7	240	12	Adi56226 Human pol
16	211.6	16.6	2177	10	Ade25693 Human CDN
17	211.6	16.6	2196	3	Aaa72428 Human nuc
18	211.6	16.6	2196	10	Adj56314 Human CDN
19	210	16.5	1716	3	Aaz50465 Human mus
20	208.6	16.4	2382	6	Aas94828 Human DNA
21	189.6	14.9	1718	3	Aaz50464 Murine mu

22	118	9.3	118	2	AAT21519	Human gen
23	83.4	6.6	963	8	ABZ18589	Group III
24	83.4	6.6	981	10	AAZ51583	Human str
25	83.4	6.6	2000	8	ADAY1938	Rice gene
c	82.6	6.5	1160	10	ADC32623	Human nov
26	82.6	6.5	1276	10	ADC30883	Human nov
27	82.6	6.5	1395	12	ADQ25304	Human sof
28	82.6	6.5	1406	8	ACC62520	Human sec
29	82.6	6.5	1406	8	ACC62491	Human sec
30	82.6	6.5	531	4	ABL16071	Drosophil
31	82.2	6.4	1004	6	ABK71581	Human dit
32	81.8	6.2	1140	6	ABQ98217	Human ost
33	78.4	6.1	1066	4	ABL08723	Drosophil
34	77.6	6.1	5101	4	ABL08722	Drosophil
c	77.6	6.1	474	5	AAS76438	DNA encod
35	77.6	6.1	609	12	ACH89545	Human gen
36	71.8	5.6	797	11	ADP65591	Human bas
37	71.8	5.6	1396	11	ADP65052	Human twi
38	71.8	5.6	1664	12	ADQ22344	Human sof
c	71.8	5.6	1800	6	ABT06509	Human sof
39	71.8	5.6	1800	11	ADM83716	Human twi
40	71.8	5.6	2870	12	ADQ17684	Human sof
41	71.6	5.6	377	12	ADQ21461	Human sof
42	71.6	5.6	450	5	AAF27265	Mouse ato

## ALIGNMENTS

RESULT 1						
AAZ57861						
ID	AAZ57861	standard;	cdNA;	1272	BP.	
XX	AC					
XX	AAZ57861;					
DT	11-APR-2000	(first entry)				
DE	Protein regulating gene expression	PRGE-23	cdNA clone	3340296.		
XX	Protein regulating gene expression;	PRGE-23;	human;	cancer;	inflammation;	
KW	anticancer;	antitumor;	antiinflammatory;	Myc;	HLH protein;	gene therapy;
KW	diagnosis;	ss.				
XX	Homo sapiens.					
XX	Key	Location/Qualifiers				
FT	CDS	268..807				
FT		/*tag= a				
XX	W09964596-A2.					
XX	16-DEC-1999.					
XX	11-JUN-1999;	99WO-US013281.				
PR	12-JUN-1998;	98US-0089029P.				
PR	29-JUL-1998;	98US-0094575P.				
PR	14-OCT-1998;	98US-0104624P.				
XX	(INCY-) INCYTE PHARM INC.					
PI	Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;					
PI	Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DAM;					
XX	WPI; 2000-116543/10.					
DR	P-PSDB; AAY58630.					
XX	New human polypeptides that regulate gene expression, for treatment,					
PT	prevention and diagnosis of, e.g. cancer.					
XX	Claim 9; Page 139; 150pp; English.					
XX	The present sequence is that of Incyte clone 3340296 encoding new human					

CC protein regulating gene expression PRGE-23 (see AY58630). The cDNA was  
 CC initially isolated from spleen tissue cDNA library SPLN010, and the  
 CC full-length sequence assembled from overlapping clones from a number of  
 CC libraries. PRGE-23 is expressed in reproductive, developmental and  
 CC urologic tissues associated with cancer, inflammation and foetal  
 CC diseases, disorders or conditions. It is characterised as an Mvz-type HLH  
 CC protein. The invention provides PRGE polypeptides (see AY58608-38) and  
 CC polynucleotides (see AAZ57839-69), expression vectors, host cells,  
 CC antibodies, agonists and antagonists. It also provides methods for  
 CC diagnosing, treating or preventing disorders associated with expression  
 CC of PRGE. Polynucleotides are also used as sources of probes and primers  
 CC for diagnosis and monitoring of disease, also for detecting related  
 CC sequences and in gene mapping  
 XX  
 SQ Sequence 1272 BP; 311 A; 358 C; 318 G; 285 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 1272; DB 3; Length 1272;  
 Best Local Similarity 100.0%; Pred. No. 5,1e-284;  
 Matches 1272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCTACGGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTGCTGCAGAGTC 60  
 DB 1 TCTACGGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTGCTGCAGAGTC 60  
 QY 61 CTCGGGGTTCCTTCTCAACACTCTGCAAGGGAAAGGTTGTGAGACCCACACAGACC 120  
 DB 61 CTCGGGGTTCCTTCTCAACACTCTGCAAGGGAAAGGTTGTGAGACCCACACAGACC 120  
 QY 121 CAACCTCCAGCTCCACAGAGGTGCTGCGCCACACTCGGAGGCGCTTGGTTTCAGG 180  
 DB 121 CAACCTCCAGCTCCACAGAGGTGCTGCGCCACACTCGGAGGCGCTTGGTTTCAGG 180  
 QY 181 GTCTCTCTGCTCTCTCTCACCCTTCTCTGCTTCTCTGCTCTCTCTCTCTCTCTCT 240  
 DB 181 GTCTCTCTGCTCTCTCTCACCCTTCTCTGCTTCTCTGCTCTCTCTCTCTCTCTCT 240  
 QY 241 CT 300  
 DB 241 CT 300  
 QY 301 CTTCAAGAGGTGAGATGTTGNAATGTGACGGGTGAAATGGATTGCAACAGGAATTT 360  
 DB 301 CTTCAAGAGGTGAGATGTTGNAATGTGACGGGTGAAATGGATTGCAACAGGAATTT 360  
 QY 361 GTGACTTCCAAACGAGAGACCGAGAGAGCTCCAACTGCGAGATGGGTCTCTCCAGAG 420  
 DB 361 GTGACTTCCAAACGAGAGACCGAGAGAGCTCCAACTGCGAGATGGGTCTCTCCAGAG 420  
 QY 421 GGCCGGGGCGCTGGGCAAG 480  
 DB 421 GGCCGGGGCGCTGGGCAAG 480  
 QY 481 GTGAGCCAG 540  
 DB 481 GTGAGCCAG 540  
 QY 541 ATGCGAGTGTGAGCAAGCGCTTCTCCAGACTCAAGACCACTGCGGTGGTGGTGGTGG 600  
 DB 541 ATGCGAGTGTGAGCAAGCGCTTCTCCAGACTCAAGACCACTGCGGTGGTGGTGGTGG 600  
 QY 601 GACACCAAGCTCTCAAGCTGGACAGCTCAGGCTGGGCTGCGCTGAGTACATGCCACATG 660  
 DB 601 GACACCAAGCTCTCAAGCTGGACAGCTCAGGCTGGGCTGCGCTGAGTACATGCCACATG 660  
 QY 661 AGGCAGATCCTGGCTAACGACAAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 DB 661 AGGCAGATCCTGGCTAACGACAAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 QY 721 TGCCCTTTATGTGGCGGGAACCCGAGAGTACCTGAAGAGAGTGGTGAACCCGAGC 780  
 DB 721 TGCCCTTTATGTGGCGGGAACCCGAGAGTACCTGAAGAGAGTGGTGAACCCGAGC 780  
 QY 781 CGCTTATGTGAACCAACCGCGTCTCTGACCTTGGAGGTGCGAGTCTGGGAAAGGCGGCTC 840

DB 781 CGCTTATGTGAACCAACCGCGTCTCTGACCTTGGAGTGGAGTCTGGAAAGGCGGCTC 840  
 QY 841 CCGGGGGAGAGCGGGCCCGGGGAAAGGAGACCCCTGCGCTCAGTGTCTCTGTCTCTCTTC 900  
 DB 841 CCGGGGGAGAGCGGGCCCGGGGAAAGGAGACCCCTGCGCTCAGTGTCTCTGTCTCTCTTC 900  
 QY 901 CCCTCGCAATGTCT 960  
 DB 901 CCCTCGCAATGTCT 960  
 QY 961 ATTCTGTTTCCAAACAGAGAGAGATCAATGTACTTACAAAGATTCCCATCTATTTAACTT 1020  
 DB 961 ATTCTGTTTCCAAACAGAGAGAGATCAATGTACTTACAAAGATTCCCATCTATTTAACTT 1020  
 QY 1021 TATTAATCTTACCGTGAATGACTCTGCAAGCCTTCTGCTGCTCCAGTGAATATGTAAT 1080  
 DB 1021 TATTAATCTTACCGTGAATGACTCTGCAAGCCTTCTGCTGCTCCAGTGAATATGTAAT 1080  
 QY 1081 ATAAATATAAATAGATAAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAT 1140  
 DB 1081 ATAAATATAAATAGATAAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAT 1140  
 QY 1141 GTAGATCATAGGATAGCTGACTTTGACAGTCAATTTATAAGTAATTCACITTAAGATA 1200  
 DB 1141 GTAGATCATAGGATAGCTGACTTTGACAGTCAATTTATAAGTAATTCACITTAAGATA 1200  
 QY 1201 TATATTTTTTCAACAAGTTTGTCTACTTTGCAAAATAACTCTTTTATATCTCTAA 1260  
 DB 1201 TATATTTTTTCAACAAGTTTGTCTACTTTGCAAAATAACTCTTTTATATCTCTAA 1260  
 QY 1261 AAAAAAAAAAAAAA 1272  
 DB 1261 AAAAAAAAAAAAAA 1272  
 RESULT 2  
 ABL65519  
 ID ABL65519 standard; DNA; 697 BP.  
 XX  
 AC ABL65519;  
 XX  
 DT 15-MAY-2002 (first entry)  
 XX  
 DE Lung cancer related gene sequence SEQ ID NO:3856.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytosstatic; Gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200194629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PP 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.







PT cells.  
XX  
PS Disclosure; Page 367; 444pp; English.  
XX

The invention relates to a method of diagnosing (I) the onset or progression of benign prostatic hyperplasia (BPH), or screening (II) for or identifying an agent that modulates the onset or progression of BPH. The method is based on changes in gene expression in BPH tissue isolated from patients exhibiting different clinical states of prostate hyperplasia as compared to normal prostate tissue. (I) comprises detecting the expression levels of one or more genes in prostate cells from the subject that are differentially regulated compared to normal prostate cells. (II) comprises preparing a first gene expression profile of BPH cells or BPH-like cell population, exposing the cells to the agent, preparing a second gene expression profile of the agent exposed cells, and comparing the first and second gene expression profiles. (I) is useful for diagnosing the onset or progression of BPH. (II) is useful for identifying an agent that modulates the onset or progression of BPH. The methods are useful to present information identifying the expression level in a tissue or cells, by comparing the expression level of genes given in the specification in the tissue or cells to the level of expression of gene in the database, and displaying the expression levels of at least one gene in the tissue or cell sample compared to the expression level in BPH. Agents using (II) are useful for treating BPH or prostate cancer. ABK64106-ABK64860 represent human benign prostatic hyperplasia gene sequences of the invention

XX  
SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;

Query Match 45.7%; Score 581; DB 6; Length 697;  
Best Local Similarity 95.1%; Pred. No. 2.8e-124;  
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

QY 620 TGGACACGCTCAGGCTGGGCTCCAGCTACATCCGCTTGGAGGAGATCCCTGGTAAAG 679  
DB 1 TGGACACGCTCAGGCTGGGCTCCAGCTACATCCGCTTGGAGGAGATCCCTGGTAAAG 60

QY 680 ACAATACGAGACGGGTACATCCCGTCAACCTGACGTGGCGCTTTATGFGGCG 739  
DB 61 ACAATACGAGACGGGTACATCCCGTCAACCTGACGTGGCGCTTTATGFGGCG 120

QY 740 GAAACCCGAGAGTACCTGAAGAGTGTGACGCGGCGCTTATGTGAACACCG 799  
DB 121 GAAACCCGAGAGTACCTGAAGAGTGTGACGCGGCGCTTATGTGAACACCG 180

QY 800 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGCTCCCGGGGGA-GCGGGCCCC 858  
DB 181 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGGCGCGCTCCCGGGGANGCGCNCCT 240

QY 859 GGAAGGCGACCCCTGCCCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 918  
DB 241 GGAAGGCGACCCCTGCCCTGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 300

QY 919 TCTCTGTCACCCCGGAGACACTTTACACGACGAGGAGTTCGTTTCCAAACCGA 978  
DB 301 TCTCTGTCACCCCGGAGACACTTTACACGACGAGGAGTTCGTTTCCAAACCGA 360

QY 979 GGAGATCAATGTACTTACAAAGATTCCTATCTTAACTTATTAATCTTACCGTGA 1038  
DB 361 GGAGATCAATGTACTTACAAAGATTCCTATCTTAACTTATTAATCTTACCGTGA 420

QY 1039 ATGACTCTCAAGCCTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1098  
DB 421 ATGACTCTCAAGCCTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480

QY 1099 AAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAGATGATAGATAGTACT 1158  
DB 481 AAGAGCCTATCAATGTATCTTTGTACAAATATGTTGTAAGATGATAGATAGTACT 540

QY 1159 GACTTTGACAGTCACATTTATTAAGTAACTTAAAGATATATATTTTTCACAA 1218  
DB 541 GACTTTGACAGTCACATTTATTAAGTAACTTAAAGATATATATTTTTCACAA 597

QY 1219 GTTTTGCTACTTTGAAATAAATCTTTCTTATATGCTTAAATATATATATATAT 1271  
DB 598 -AGTTGCACTTTTGAATAAATCTTTCTTATATGCTTAAATATATATATATAT 649

RESULT 5  
ABQ55427  
ID ABQ55427 standard; cDNA; 632 BP.  
XX  
AC ABQ55427;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Human ovarian antigen HNOJ185 cDNA, SEQ ID NO:1307.  
XX  
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200200677-A1.  
XX  
PD 03-JAN-2002.  
XX  
PF 07-JUN-2001; 2001WO-US018569.  
XX  
PP 07-JUN-2000; 2000US-0209467P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Birse CE, Rosen CA;  
XX  
XX WPI; 2002-147878/19.  
DR P-PSDB; ABP42350.  
XX  
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
XX useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
XX cancer), immune disorders, cardiovascular disorders and neurological  
XX diseases.  
PS Claim 1; SEQ ID NO 1307; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-  
ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
encompasses polypeptides 90% identical and polynucleotides 95% identical  
to the sequences of the invention. The invention additionally relates to  
recombinant vectors and host cells comprising human ovarian antigen  
polynucleotides, antibodies against human ovarian antigens, and the use  
of ovarian antigen polynucleotides and polypeptides in diagnosing,  
treating, prognosing or preventing various ovary and/or breast-related  
disorders. Such conditions include ovarian cancer and breast cancer, and  
metastatic tumours of ovarian or breast origin, reproductive system  
disorders (e.g., infertility, disorders of pregnancy, anovulation,  
polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
vaginitis), immune disorders (e.g., congenital and acquired  
immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
blood-related disorders (e.g., anaemia), cardiovascular disorders,  
respiratory disorders, neurological disorders, gastrointestinal disorders  
and urinary system disorders. Ovarian antigen polypeptides and  
polynucleotides may also be used in screening for compounds which  
modulate ovarian antigen expression or activity. The polynucleotides may  
further be used for gene therapy, chromosome mapping, in the  
identification of individuals and in forensic analysis, and the

CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX

SQ Sequence 632 BP; 145 A; 176 C; 201 G; 93 T; 0 U; 17 Other;

Query Match 41.3%; Score 525.6; DB 6; Length 632;  
 Best Local Similarity 93.4%; Pred. No. 1.8e-111;  
 Matches 593; Conservative 0; Mismatches 36; Indels 6; Gaps 5;

QY 257 CTCCTCCCAACATGTCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGA 316  
 Db 2 CTCCTCCCAACATGTCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGA 61  
 QY 317 TGTGGAATGTCACCGGTTGAAATGATTCGAAACAGGAATTTGTGACTTCCAAACGAGA 376  
 Db 62 TGTGGAATGTCACCGGTTGAAATGATTCGAAACAGGAATTTGTGACTTCCAAACGAGA 121  
 QY 377 GCACCGAGAGAGCTCAACTGGAGATGGGTCTCCCGAGAGGCGCGCGCGCTGG 436  
 Db 122 GCACCGAGAGAGCTCAACTGGAGATGGGTCTCCCGAGAGGCGCGCGCGCTGG 181  
 QY 437 GCAAGAGAGAGAGGCGCGCCACCAAGAGAGCGCCCTGAGCGGGTTCAGCCAGGAGGGA 496  
 Db 182 GCAAGAGAGAGAGGCGCGCCACCAAGAGAGCGCCCTGAGCGGGTTCAGCCAGGAGGGA 241  
 QY 497 AGCAGTGCAGCGCAACCGCGCCCAACCGCGAGAGCGGCGCGCATGGAGTGTGAGA 556  
 Db 242 AGCAGTGCAGCGCAACCGCGCCCAACCGCGAGAGCGGCGCGCATGGAGTGTGAGA 301  
 QY 557 AGSCCTTCTCCAGACTCAAGACACCCCTGCTGGTGCCTCCCGACACCAAGCTCTCCA 616  
 Db 302 AGSCCTTCTCCAGACTCAAGACACCCCTGCTGGTGCCTCCCGACACCAAGCTCTCCA 361  
 QY 617 AGCTGACACGCTCAGGCTGGGCTCAGTATACATCGCCACCTTGGAGGAGATCTCTGGTA 676  
 Db 362 AGCTGACACGCTCAGGCTGGGCTCAGTATACATCGCCACCTTGGAGGAGATCTCTGGTA 421  
 QY 677 ACACAAATACGAGAGCGGTACATCAACCGGTCACTGACCTGACCTGACCTTATGTGG 736  
 Db 422 ACACAAATACGAGAGCGGTACATCAACCGGTCACTGACCTGACCTTATGTGG 479  
 QY 737 CCGGGAACCCGAGAGTACCTGAAAGAGTGG-TGACCGGAGCGCGCTTATGTGGAAAC 795  
 Db 480 -CGGGAACCCGAGAGTACCTGAAAGAGTGGTTGACCGGAGCGCGTATGTGGAAAC 538  
 QY 796 ACCGCTCTGACCTTGGAGGTGCGAGTCTGGAAAGGCGCGCTCCCGGGGGAGCGGCG 855  
 Db 539 ACCGCTCTGACCTTGGAGGTGCGAGTCTGGAAAGGCGCGCTCCCGGGGGAGCGGCG 596  
 QY 856 CCGGGAACCCGAGAGTACCTGAAAGAGTGGTGGTCTCTCTG 890  
 Db 597 CCGGGAACCCGAGAGTACCTGAAAGAGTGGTGGTCTCTCTG 631

RESULT 6  
 ABL66121/c  
 ID ABL66121 standard; DNA; 471 BP.  
 XX  
 AC ABL66121;  
 XX  
 DT 15-MAY-2002 (first entry)  
 DE Lung cancer related gene sequence SEQ ID NO:4458.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX

OS Homo sapiens.  
 XX WO200194629-A2.  
 XX 13-DEC-2001.  
 XX  
 XX 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233167P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVAL-) AVALON PHARM.  
 XX  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 XX WPI; 2002-188264/24.  
 XX

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and determining a change in expression of a gene of a signature gene set.  
 Claim 1; SEQ ID NO 4458; 44pp; English.  
 CC The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is

Sequence 471 BP; 142 A; 89 C;

Sequence 471 BP: 142 A; 89 C; 110 G; 130 T; 0 U; 0 Other;





SQ Sequence 482 BP; 99 A; 156 C; 143 G; 83 T; 0 U; 1 Other;

Query Match 23.4%; Score 297.2; DB 9; Length 482;  
Best Local Similarity 97.1%; Pred. No. 1.1e-58;  
Matches 302; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 517 GCCAAGCGGAGGAGCGGCGCCGATGCTGAGTGGTCTGAGCAAGCCCTTCTCCAGACTCAAG 576  
DB 33 GCNTAGCGGAGAGGCGGCGCCGATGCTGAGTGGTCTTCCAGACTCAAG 92

QY 577 ACCACCTGCTGCTGGTGGGCGCCGACACCAAGCTTCCAGTGCACAGCTCAGGCTG 636  
DB 93 ACCACCTGCTGCTGGTGGGCGCCGACACCAAGCTTCCAGTGCACAGCTCAGGCTG 152

QY 637 GCGTCCAGCTACATCCGCCACTTGGCGAGATCTGGTAAACGACAAATACAGAACCGG 696  
DB 153 GCGTCCAGCTACATCCGCCACTTGGCGAGATCTGGTAAACGACAAATACAGAACCGG 212

QY 697 TACATTACCCCGTCAACCTGACGTGGCCCTTTATGGTGGCGGGAACCCGAGAGTGAC 756  
DB 213 TACATTACCCCGTCAACCTGACGTGGCCCTTTATGGTGGCGGGAACCCGAGAGTGAC 272

QY 757 CTGAAGAAGTGTGACCGGAGCCGCTTATGTGNAACCAACCGCTCTGACCTTGGAGG 816  
DB 273 CTGAAGAAGTGTGACCGGAGCCGCTTATGTGNAACCAACCGCTCTGACCTTGGAGT 332

QY 817 TGGGAGTCTGG 827  
DB 333 TTGGTACCTGG 343

RESULT 11  
AAC08260  
ID AAC08260 standard; cDNA; 271 BP.

AC AAC08260;  
XX  
DT 06-OCT-2000 (first entry)  
DE Human secreted protein 5' EST, SEQ ID NO: 12335.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-00200610.  
XX  
PR 26-FEB-1999; 99US-0122487P.  
XX  
PA (GIST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
DR  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX  
PS Claim 1; SEQ ID NO 12335; 71pp + Sequence Listing; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors

XX  
SQ Sequence 271 BP; 41 A; 97 C; 65 G; 67 T; 0 U; 1 Other;

Query Match 19.4%; Score 246.8; DB 3; Length 271;  
Best Local Similarity 98.4%; Pred. No. 3.9e-47;  
Matches 248; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTACGCCACGACTCTGGAGTGGGGAACAGAGAGCGGTTCTCTGCTGCAGAGTC 60  
DB 19 TCTACGCCACGACTCTGGAGTGGGGAACAGAGAGCGGTTCTCTGCTGCAGAGTC 78

QY 61 CTCGGGTTCTTCTCTCAAACTCTGCGAAGGGGAAAGGTTGTGAGACCCACACACCC 120  
DB 79 CTCGGGTTCTTCTCTCAAACTCTGCGAAGGGGAAAGGTTGTGAGACCCACACACCC 138

QY 121 CAATCTCAGTCTCCAGCAGAGAGTGGCTGCGCCACACTCGGAGGCTCTTGGTTTCAGG 180  
DB 139 CAATCTCAGTCTCCAGCAGAGAGTGGCTGCGCCACACTCGGAGGCTCTTGGTTTCAGG 198

QY 181 GTCTCTCTGTCTCTCTCACCCCTCTTCTCTGCTGCTTCTCTGCTCTCTCTCTCT 240  
DB 199 GTCTCTCTGTCTCTCTCACCCCTCTTCTCTGCTGCTTCTCTGCTCTCTCTCTCTCA 258

QY 241 CTCTCTCCCTCG 252  
DB 259 CCTCTTCTCTCG 270

RESULT 12  
ADK61063/C  
ID ADK61063 standard; DNA; 265 BP.

XX  
AC ADK61063;  
XX  
DT 06-MAY-2004 (first entry)  
DE Ovarian cancer-related DNA #218 with altered ovarian cancer expression.  
XX  
XX ds; gene; ovarian tumor; BRCA-1-like; BRCA-2-like; non-BRCA-like;  
KW gene expression; primer; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO2003068054-A2.  
XX  
PD 21-AUG-2003.  
XX  
PF 13-FEB-2003; 2003WO-US004688.  
XX  
PR 13-FEB-2002; 2002US-0357031P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
PI Jazaeri AA, Boyd J, Liu ET;  
XX WPI; 2003-689589/65.  
DR  
XX  
XX Classifying an ovarian tumor as a BRCA-1-like or BRCA-2-like or non-BRCA-  
PT like tumor by determining a pattern of expression in the ovarian tumor of  
PT several markers.  
XX  
XX Disclosure; SEQ ID NO 233; 137pp; English.  
PS  
XX The invention relates to a method of classifying an ovarian tumor as a  
CC BRCA-1-like or BRCA-2-like or non-BRCA-like tumor by: (1) determining a



CC pattern of expression in the ovarian tumor of several markers given in  
 CC the specification; and (2) comparing a similarity of the pattern of  
 CC expression of the markers in the ovarian tumor to a pattern of expression  
 CC of the markers in a comparison tissue of a known BRCA-1-like or BRCA-2-  
 CC like or non-BRCA-like tumor. The method is useful for classifying an  
 CC ovarian tumor as a BRCA-1-like or BRCA-2-like or non-BRCA-like tumor.  
 CC This sequence corresponds to an ovarian cancer -related gene having an  
 CC altered pattern of expression in ovarian cancer. (Note: The sequence data  
 CC for this patent did not form part of the printed specification but was  
 CC obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences).

XX Sequence 265 BP; 98 A; 33 C; 38 G; 96 T; 0 U; 0 Other;

Query Match 18.6%; Score 236.2; DB 10; Length 265;  
 Best Local Similarity 98.1%; Pred. No. 1.1e-44;  
 Matches 260; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 995 TACAAGATCCCATCTATTAACTTTTAACTTCTACCGTG-AATGACTCTGCAAGCC 1053  
 DB 265 TACAAGATCCCATCTATTAACTTTTAACTTCTACCGTGAAATGACTCTGCAAGCC 206  
 QY 1054 TTGCTGCTCCCAAGTCAATATGTAATATAATATAATAGATAAGAGCCTATCAATG 1113  
 DB 205 TTGCTGGACCAAGTCAATATGTAATATAATATAATAGATAAGAGCCTATCAATG 146  
 QY 1114 TATCTTTTGTACAATATGTTGAAATGTAGATCATGATAGTACGTTTGACAGTCAC 1173  
 DB 145 TATCTTTTGTACAATATGTTGAAATGTAGATCATGATAGTACGTTTGACAGTCAC 86  
 QY 1174 ATTATAAGATTAATTCACCTTAAGATATATATTTTTCACCAAGT-TTTGCTACTTTT 1232  
 DB 85 ATTATAAGATTAATTCACCTTAAGATATATATTTTTCACCAAGTATTGCTACTTTT 26  
 QY 1233 GAAATAAATCTTCTTTTATATGTC 1257  
 DB 25 GAAATAAATCTTCTTTTATATGTC 1

# RESULT 13

ADM32307/c  
 ID ADM32307 standard; DNA; 265 BP.

XX AC ADM32307;

XX 20-MAY-2004 (first entry)

XX Human cancer gene, SEQ ID NO 2.

XX detection; cancer; cytotoxic; anticancer; cytostatic; gene therapy; ds;  
 KW human.

XX Homo sapiens.

XX JP2004049122-A.

XX 19-FEB-2004.

XX 19-JUL-2002; 2002JP-00211830.

XX 19-JUL-2002; 2002JP-00211830.

XX (CHUS) CHUGAI PHARM CO LTD.

XX (GANK-) ZH GAN KENYUKAI.

XX WPI; 2004-232131/22.

XX Detecting cancer, comprises measuring the expression level of cancer  
 PT related polypeptides selected from a set of polypeptides given in the  
 PT specification.

XX Claim 1; SEQ ID NO 2; 116pp; Japanese.

XX

CC The invention relates to a novel method for detecting cancer. The novel  
 CC method involves measuring the expression level of a polypeptide encoded  
 CC by a DNA having a sequence chosen from 119 fully defined sequences such  
 CC as 416, 265, 450, 376, 360, 245, 290, 288, 221, 391 nucleotides etc., as  
 CC given in the specification, or a DNA that hybridizes to the above DNA  
 CC under stringent conditions. The invention further relates to: a  
 CC diagnostic test of cancer; a DNA which encodes a cytotoxic protein and is  
 CC coupled with a functional promoter region of a cancer detecting DNA; a  
 CC vector having cytotoxic DNA inserted in it; a transformed cell containing  
 CC the said vector; a method for evaluating the presence or absence of  
 CC anticancer activity in a test sample; and producing a composition, by  
 CC mixing a sample evaluated by the anticancer activity detection method and  
 CC a carrier. The novel polynucleotides have cytostatic activity. The vector  
 CC containing a cytotoxic polynucleotide can be used to treat cancer by gene  
 CC therapy. The novel method of the invention allows early detection of  
 CC cancer. This polynucleotide represents one of the 119 DNA sequences  
 CC useful in the method for detecting cancer of the invention.

XX Sequence 265 BP; 98 A; 33 C; 38 G; 96 T; 0 U; 0 Other;

Query Match 18.6%; Score 236.2; DB 12; Length 265;  
 Best Local Similarity 98.1%; Pred. No. 1.1e-44;  
 Matches 260; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 995 TACAAGATCCCATCTATTAACTTTTAACTTCTACCGTG-AATGACTCTGCAAGCC 1053  
 DB 265 TACAAGATCCCATCTATTAACTTTTAACTTCTACCGTGAAATGACTCTGCAAGCC 206  
 QY 1054 TTGCTGCTCCCAAGTCAATATGTAATATAATATAATAGATAAGAGCCTATCAATG 1113  
 DB 205 TTGCTGGACCAAGTCAATATGTAATATAATATAATAGATAAGAGCCTATCAATG 146  
 QY 1114 TATCTTTTGTACAATATGTTGAAATGTAGATCATGATAGTACGTTTGACAGTCAC 1173  
 DB 145 TATCTTTTGTACAATATGTTGAAATGTAGATCATGATAGTACGTTTGACAGTCAC 86  
 QY 1174 ATTATAAGATTAATTCACCTTAAGATATATATTTTTCACCAAGT-TTTGCTACTTTT 1232  
 DB 85 ATTATAAGATTAATTCACCTTAAGATATATATTTTTCACCAAGTATTGCTACTTTT 26  
 QY 1233 GAAATAAATCTTCTTTTATATGTC 1257  
 DB 25 GAAATAAATCTTCTTTTATATGTC 1

# RESULT 14

ACA56430  
 ID ACA56430 standard; cDNA; 240 BP.

XX AC ACA56430;

XX 06-JUN-2003 (first entry)

XX Chicken signalling pathway polynucleotide probe SEQ ID NO 1028.

XX Chicken; probe; ss; array element; Parkinson's disease;  
 KW signalling pathway population; cancer; adenocarcinoma; leukaemia;  
 KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.

XX Gallus gallus.

XX US6500938-B1.

XX 31-DEC-2002.

XX 30-JAN-1998; 98US-00016434.

XX 30-JAN-1998; 98US-00016434.

XX (INCY-) INCYTE GENOMICS INC.

XX Au-Young J, Seilhamer JJ;

XX

1026c and cdp



Result No.	Score	Query Match	Length	DB	ID	Description
1	246.8	19.4	271	4	US-09-513-999C-12335	Sequence 12335, A
2	213	16.7	240	4	US-09-016-434-1028	Sequence 1028, Ap
3	71.8	5.6	1800	4	US-09-771-357-106	Sequence 106, App
C 4	64.6	5.1	2397	1	US-07-891-942G-1	Sequence 11, Appl
C 5	63.4	5.0	305	1	US-08-253-155A-4	Sequence 4, Appl
6	63.4	5.0	1268	1	US-08-910-973-12	Sequence 12, Appl
7	63.4	5.0	1268	4	US-09-499-227-12	Sequence 12, Appl
8	63.2	5.0	3618	3	US-09-042-353-224	Sequence 224, App
9	63.2	5.0	3618	3	US-08-758-417A-72	Sequence 72, Appl
10	63.2	5.0	3698	1	US-07-834-533A-49	Sequence 49, Appl
11	63.2	5.0	3698	2	US-08-800-353-49	Sequence 49, Appl
12	63.2	5.0	3698	5	PCI-US92-06185-49	Sequence 49, Appl
13	63.2	5.0	3699	1	US-08-053-131-120	Sequence 120, App
14	63.2	5.0	3699	1	US-08-645-641-120	Sequence 120, App
15	63.2	5.0	3699	1	US-07-853-408B-120	Sequence 120, App
16	63.2	5.0	3699	2	US-08-096-763-120	Sequence 120, App
17	63.2	5.0	3699	2	US-08-308-863-120	Sequence 120, App
18	63.2	5.0	3699	5	PCI-US92-10983-120	Sequence 120, App
19	63	5.0	1352	1	US-08-552-142A-10	Sequence 10, Appl
20	63	5.0	1535	1	US-08-910-973-10	Sequence 10, Appl
21	63	5.0	1535	4	US-09-499-227-10	Sequence 10, Appl
22	63	5.0	1550	3	US-09-234-332-3	Sequence 3, Appl
C 23	62.6	4.9	349	1	US-08-253-155A-14	Sequence 14, Appl
24	62.2	4.9	165	4	US-09-636-215-736	Sequence 736, App
25	62.2	4.9	165	4	US-09-685-166A-736	Sequence 736, App
26	62.2	4.9	165	4	US-09-679-426-736	Sequence 736, App
27	62	4.9	310	1	US-08-552-142A-12	Sequence 12, Appl





```
; CLONE: 20A1 (neuroD3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
US-08-910-973-12

Query Match
Best Local Similarity 5.0%; Score 63.4; DB 1; Length 1268;
Matches 103; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 504 CAGCGCAACGCGCCCAACGCGAGAGCGGGCCCGCATGCGAGTGTGAGCAAGGCCTT 563
Db 330 CCGCGCGGTCAAGGCCAACGATCGCGAGCGCAACCGCATGCAAACTTGAACGCGGCCT 389
QY 564 CTCGAGACTCAAGACCAACCTGCGCTGGTGGCCCGCCGACCAAGCTCTCCAACTGGA 623
Db 390 GGACGACTGCGGAGCGTGTGCGCTGCTTCCCGAGACACCAAGCTCACCAAAATCGA 449
QY 624 CAGCTCAGCTGGCTCCAGCTACATCGCCCACTTGGGCGAGATCCTG 672
Db 450 GAGCTGCGCTTGGCTTACACTACATCTGGGCTCTGGCGGAGACACTG 498

RESULT 7
US-09-499-227-12
; Sequence 12, Application US/09499227
; Patent No. 644463
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/499,227
; FILING DATE: 05-August-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-May-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-May-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,973
; FILING DATE: 07-August-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-12742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1268 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: 20A1 (neuroD3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
US-09-499-227-12

Query Match
Best Local Similarity 5.0%; Score 63.4; DB 4; Length 1268;
Matches 103; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 504 CAGCGCAACGCGCCCAACGCGAGAGCGGGCCCGCATGCGAGTGTGAGCAAGGCCTT 563
Db 330 CCGCGCGGTCAAGGCCAACGATCGCGAGCGCAACCGCATGCAAACTTGAACGCGGCCT 389
QY 564 CTCGAGACTCAAGACCAACCTGCGCTGGTGGCCCGCCGACCAAGCTCTCCAACTGGA 623
Db 390 GGACGACTGCGGAGCGTGTGCGCTGCTTCCCGAGACACCAAGCTCACCAAAATCGA 449
QY 624 CAGCTCAGCTGGCTCCAGCTACATCGCCCACTTGGGCGAGATCCTG 672
Db 450 GAGCTGCGCTTGGCTTACAACTACATCTGGGCTCTGGCGGAGACACTG 498

RESULT 8
US-09-042-353-224
; Sequence 224, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,353
; FILING DATE: 13-Mar-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-Mar-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: WO PCT/US96/16433  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/758,417  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/21803  
FILING DATE: 01-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 014643-009040US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 224:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3618 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-09-042-353-224

Query Match 5.0%; Score 63.2; DB 3; Length 3618;  
Best Local Similarity 67.4%; Pred. No. 9e-06;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 144 TGGCTCGCCACACTCGGAGGCTCTTGGTTTCAGGGTCTCTGTCTCTCTCTCACC 203  
DB 152 TGTGTGTC 211  
QY 204 TCTTCTCGCTTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263  
DB 212 TGTCTGTC 271  
QY 264 AACATGTCCAC 275  
DB 272 ACACACACAC 283

RESULT 9  
US-08-758-417A-72  
Sequence 72, Application US/08758417A  
Patent No. 6300129  
GENERAL INFORMATION:  
APPLICANT: Lonberg, Nils  
Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for  
Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 417  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758,417A  
FILING DATE: 02-DEC-1996  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995  
APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Serafini, Andrew T.  
REGISTRATION NUMBER: 41,303  
REFERENCE/DOCKET NUMBER: 014643-009030US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 72:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3618 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..3618  
OTHER INFORMATION: /note= "vector pGPe"  
SEQUENCE DESCRIPTION: SEQ ID NO: 72:  
US-08-758-417A-72

Query Match 5.0%; Score 63.2; DB 3; Length 3618;  
Best Local Similarity 67.4%; Pred. No. 9e-06;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 144 TGGCTCGCCACACTCGGAGGCTCTTGGTTTCAGGGTCTCTGTCTCTCTCTCACC 203  
DB 152 TGTGTGTC 211  
QY 204 TCTTCTCGCTTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263  
DB 212 TGTCTGTC 271  
QY 264 AACATGTCCAC 275  
DB 272 ACACACACAC 283

RESULT 10





```

; REGISTRATION NUMBER: 87654
; REFERENCE/DOCKET NUMBER: 14643-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3698 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US92-06185-49

Query Match          5.0%; Score 63.2; DB 5; Length 3698;
Best Local Similarity 67.4%; Pred.No.9.1e-06;
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 20;

QY      144 TGGCTGCCACACATCGGAGGCTCTGTGTTTCAGGGTCTCTCTCTCTCTCTCTCTCTC 203
Db      153 TGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 212
QY      204 TCTTCTCGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 263
Db      213 TGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 272
QY      264 AAACATGTCCAC 275
Db      273 ACACACACAC 284

RESULT 13
US-08-053-131-120
; Sequence 120, Application US/08053131
; Patent No. 5661016
; GENERAL INFORMATION:
; APPLICANT: Lomberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 200
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/053,131
; FILING DATE: 26-APR-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 14643-9-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422

```

Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 144 TGGCTGGCCACACTGGGAGGCTCTTGGTTTCAGGGTCTCTGTCTCTCTCTCACCC 203  
Db 153 TGGTGTGTC 212  
QY 204 TCTTCCTCGCTTC 263  
Db 213 TGTCTCTGTC 272  
QY 264 AACATGTCCAC 275  
Db 273 ACACACACAC 284

## RESULT 15

US-07-853-408B-120  
; Sequence 120, Application US/07853408B  
; Patent No. 5789650  
; GENERAL INFORMATION:  
; APPLICANT: Lonberg, Nils  
; APPLICANT: Kay, Robert M.  
; TITLE OF INVENTION: Transgenic No. 5789650-Human Animals for  
; TITLE OF INVENTION: Producing Heterologous Antibodies  
; NUMBER OF SEQUENCES: 150  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: William M. Smith  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/853,408B  
; FILING DATE: 19920318  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M.  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 14643-9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 120:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3699 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-07-853-408B-120

Query Match 5.0%; Score 63.2; DB 1; Length 3699;  
Best Local Similarity 67.4%; Pred. No. 9.1e-06;  
Matches 89; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 144 TGGCTGGCCACACTGGGAGGCTCTTGGTTTCAGGGTCTCTGTCTCTCTCTCACCC 203  
Db 153 TGGTGTGTC 212  
QY 204 TCTTCCTCGCTTC 263  
Db 213 TGTCTCTGTC 272  
QY 264 AACATGTCCAC 275  
Db 273 ACACACACAC 284

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 00:04:02 ; Search time 455 Seconds

(without alignments)  
2065.157 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179

Sequence: 1 MSTGSLSDVDLQEVMELEC.....PESDLKEYVTASRLCGTTAS 179

Scoring table:

OLIGO 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame\_plus\_p2n.model  
-Q=/cgn2.1/USFTO spool\_h/US0701674/runat\_15122004\_100529\_28299/app\_query.fasta\_1.327  
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0701674@cgn\_1.1.470@runat\_15122004\_100529\_28299 -NCFU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_23Sep04:\*

1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179	100.0	1272	3	Aaz57861 Protein r
2	94	52.5	482	3	Ach21857 Human adu
3	61	34.1	697	6	Ab165519 Lung canc
4	61	34.1	697	6	Ab165935 Lung canc
5	61	34.1	697	6	Abk64784 Human ben
6	49	27.4	632	6	Abq55427 Human ova

7	36	20.1	240	10	ACA56430	Chicken s
8	36	20.1	240	12	ADJ56226	Human pol
9	28	15.6	493	9	ACH23369	Human adu
10	28	15.6	1716	3	AAZ50465	Human mus
11	28	15.6	1718	3	AAZ50464	Murine mu
12	28	15.6	2177	10	ADZ52593	Human cDN
13	28	15.6	2196	3	AAA72428	Human nuc
14	28	15.6	2196	10	ADJ56314	Human cDN
15	28	15.6	2382	6	AAS94828	Human DNA
16	18	10.1	85	6	ABN27386	Rat splic
17	15	8.4	1086	4	ABL08723	Drosophil
18	15	8.4	5101	4	ABL08722	Drosophil
19	12	6.7	963	8	ABZ18589	Group III
20	11	6.1	474	5	AAS76438	DNA encod
21	11	6.1	966	5	AAS76439	DNA encod
22	9	5.0	730	4	AAS12517	Gene #10
23	9	5.0	730	10	ABQ77254	Human neu
24	9	5.0	981	10	ABQ77254	Human neu
25	9	5.0	1140	6	ABQ88217	Human ost
26	9	5.0	2610	11	ACH95595	Klebsiell
27	9	5.0	2787	4	ABL10299	Drosophil
28	9	5.0	4852	4	ABL10298	Drosophil
29	9	5.0	6288	5	AAS45090	cDNAenco
30	9	5.0	12905	3	AAS39101	S. fradia
31	9	5.0	12905	3	AAS39100	S. fradia
32	8	4.5	276	12	ADQ07882	Soybean e
33	8	4.5	302	8	ABX55437	Bovine ES
34	8	4.5	353	10	ADK57797	Plant DNA
35	8	4.5	375	10	AB241346	N. gonorr
36	8	4.5	375	10	AB240186	N. gonorr
37	8	4.5	378	3	AAA75746	DNA encod
38	8	4.5	381	8	ABX09886	N. mening
39	8	4.5	405	8	ABX55489	Bovine ES
40	8	4.5	420	3	AAZ54169	Neisseria
41	8	4.5	420	3	AAZ54168	Neisseria
42	8	4.5	420	3	AAZ54170	Neisseria
43	8	4.5	423	10	AB240195	N. gonorr
44	8	4.5	423	10	AB241348	N. gonorr
45	8	4.5	518	6	ABQ52056	Oligonuc

## ALIGNMENTS

RESULT 1

AZ57861

ID AZ57861 standard; cDNA; 1272 BP.

XX AC AZ57861;

XX AC AZ57861;

DT 11-APR-2000 (first entry)

XX DE Protein regulating gene expression PRGE-23 cDNA clone 3340296.

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX KW anticancer; antitumour; antiinflammatory; Myc; HLH protein; gene therapy;

XX KW diagnosis; ss.

XX KW Homo sapiens.

XX OS Homo sapiens.

XX FH Key

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

XX FT CDS

(INCY-) INCYTE PHARM INC.

PA Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;  
 PI Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DM;  
 XX  
 XX WPI; 2000-116543/10.  
 DR P-PSDB; AAY58630.  
 XX  
 XX New human polypeptides that regulate gene expression, for treatment,  
 PT prevention and diagnosis of, e.g. cancer.  
 XX  
 XX Claim 9; Page 139; 150pp; English.

XX The present sequence is that of Incyte clone 3340296 encoding new human  
 CC protein regulating gene expression PRGE-23 (see AAY58630). The cDNA was  
 CC initially isolated from spleen tissue cDNA library SPLN0T10, and the  
 CC full-length sequence assembled from overlapping clones from a number of  
 CC libraries. PRGE-23 is expressed in reproductive, developmental and  
 CC urologic tissues associated with cancer, inflammation and foetal  
 CC diseases, disorders or conditions. It is characterised as an Myc-type HLH  
 CC protein. The invention provides PRGE polypeptides (see AAY58608-38) and  
 CC polynucleotides (see AAY57839-69), expression vectors, host cells,  
 CC antibodies, agonists and antagonists. It also provides methods for  
 CC diagnosing, treating or preventing disorders associated with expression  
 CC of PRGE. Polynucleotides are also used as sources of probes and primers  
 CC for diagnosis and monitoring of disease, also for detecting related  
 CC sequences and in gene mapping

XX  
 SQ Sequence 1272 BP; 311 A; 358 C; 318 G; 285 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2,118-172 Length: 1272  
 Score: 179.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0

US-09-701-674A-23 (1-179) x AAY57861 (1-1272)

QY 1 MetSerThrGlySerIleuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 268 ATGTCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGATGT 327

QY 21 AspGlyLeuIleuMetAspSerAsnIleuValThrSerAsnGluSerThrGluGlu 40  
 DB 328 GACGGGTTGAAATGATTCGAACAGGAATTTGACTTCAACGAGAGCAGCCAGGAG 387

QY 41 SerSerAsnGlyGluAsnGlySerProGlnIleuGlyArgGlyGlyLeuGlyLysArgArg 60  
 DB 388 AGCTCCAACTCGAGAGTAATGGGTCTCCCAAGAGGCGCGCGGCTCGGCAAGAGGAGG 447

QY 61 LysAlaProThrIleuSerProIleuSerGlyValSerGlnGluGlyValGlnValCln 80  
 DB 448 AAGGGCGCCACCAAGAGAGCCCTCAGCGGGGTCAAGAGGAGGAGGAGGAGGAGGAG 507

QY 81 ArgAsnAlaAlaAsnAlaGluArgAlaArgMetArgValLeuSerIleuAlaPheSer 100  
 DB 508 CGCAACGCGCCACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 567

QY 101 ArgLeuIleuThrThrLeuProTrpValProProAspThrLysLeuSerIleuAspThr 120  
 DB 568 AGACTCAAGACCAACCCCTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 627

QY 121 LeuArgLeuAlaSerSerTrpIleuAlaHisLeuArgGlnIleuLeuAlaAsnAspLysTrp 140  
 DB 628 CTCAGGCTGGGCTCCAGCTACATCCGCCCTTTAGGCGAGATCCTGGCTAACCAATAC 687

QY 141 GluAsnGlyTrpIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 DB 688 GAGACGGGTATCATTCACCGGTCAACCTGAGCTGGGCTTTATGGTGGCGGGAACCC 747

QY 161 GluSerAspLeuLysGluValThrAlaSerArgLeuGlyThrThrAlaSer 179

DB 748 GAGAGTGACCTGAAGAGAGTGTGACCGGAGCGCTTATGTGGAACCAACCGCGTCC 804

RESULT 2  
 ACH21857  
 ID ACH21857 standard; cDNA; 482 BP.  
 XX AC ACH21857;  
 XX DT 13-OCT-2003 (first entry)  
 XX AC Human adult ovary cDNA #237.  
 XX DE Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 XX KW genome mapping; biodiversity; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN US2003073623-A1.  
 XX PD 17-APR-2003.  
 XX PF 30-JUL-2001; 2001US-00918995.  
 XX PR 30-JUL-2001; 2001US-00918995.  
 XX PA (DRMA/) DRMANAC R T.  
 XX PA (LABA/) LABAT I.  
 XX PA (STAC/) STACHE-CRAIN B.  
 XX PA (DICK/) DICKSON M C.  
 XX PA (JONE/) JONES L W.  
 XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
 XX WPI; 2003-615964/58.  
 XX New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.

XX Claim 1; SEQ ID NO 9069; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH21857-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX SQ Sequence 482 BP; 99 A; 156 C; 143 G; 83 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 3,458-86 Length: 482  
 Score: 94.00 Matches: 94  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x ACH21857 (1-482)

```

QY 86 AlaArgGluArgAlaArgVetArgValLeuSerLysAlaPheSerArgLeuLysThrThr 105
Db 39 GCGGAGAGCGGCGCGATGGAGTGTGAGCAAGCGCTTCTCCAGACTCAAGACCCACC 98
QY 106 LeuProTirValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
Db 99 CTGCGCTGGTGGTGGCGCCCGACACCAAGCTCTCAAGCTGGACACGCTCAGGCTGGCGTCC 158
QY 126 SerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIle 145
Db 159 AGTACATCGCCCACTTGAGGCGAGATCTCTGGCTAAGCAAAATACGAGAACGGGTACATT 218
QY 145 HisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAspLysLys 165
Db 219 CACCGCGTCAACTGACGTGGCGCCCTTATGGTGGCGGGAACCCGAGAGTGACCTGAAA 278
QY 166 GluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 279 GAAGTGGTGACCGGAGCGCGCTTATGTGGAAACCCCGCGTCC 320

RESULT 3
ABL65519
ID ABL65519 standard; DNA; 697 BP.
XX
XX
AC ABL65519;
XX
XX
15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:3856.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233161P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 26-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.

```

```

PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
PS Claim 1; SEQ ID NO 3856; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;

Alignment Scores:
Pred. No.: 2.06e-52 Length: 697
Score: 61.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 34.08% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABL65519 (1-697)
QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138
Db 3 GACACGCTCAGGCTGGCGTCCAGCTCATCGCCCACTTGAGGCGAGATCTCGGTACGAC 62
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158
Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGACCTGGCCCTTTATGGTGGCGGG 122
QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178
Db 123 AAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGCGCTTATGTGGAAACCCCGCG 182
QY 179 Ser 179
Db 183 TCC 185

RESULT 4

```

ABL65935  
ID ABL65935 standard; DNA; 697 BP.  
XX AC ABL65935;  
XX  
XX 15-MAY-2002 (first entry)  
XX  
XX Lung cancer related gene sequence SEQ ID NO:4272.  
XX  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200194629-A2.  
XX  
XX 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US010838.  
XX  
XX 05-JUN-2000; 2000US-0209473P.  
XX  
XX 05-JUN-2000; 2000US-0209531P.  
XX  
XX 18-SEP-2000; 2000US-0233133P.  
XX  
XX 18-SEP-2000; 2000US-0233617P.  
XX  
XX 20-SEP-2000; 2000US-0234009P.  
XX  
XX 20-SEP-2000; 2000US-0234034P.  
XX  
XX 20-SEP-2000; 2000US-0234052P.  
XX  
XX 22-SEP-2000; 2000US-0234509P.  
XX  
XX 22-SEP-2000; 2000US-0234567P.  
XX  
XX 25-SEP-2000; 2000US-0234923P.  
XX  
XX 25-SEP-2000; 2000US-0234924P.  
XX  
XX 25-SEP-2000; 2000US-0235077P.  
XX  
XX 25-SEP-2000; 2000US-0235082P.  
XX  
XX 25-SEP-2000; 2000US-0235134P.  
XX  
XX 25-SEP-2000; 2000US-0235280P.  
XX  
XX 26-SEP-2000; 2000US-0235637P.  
XX  
XX 26-SEP-2000; 2000US-0235638P.  
XX  
XX 27-SEP-2000; 2000US-0235711P.  
XX  
XX 27-SEP-2000; 2000US-0235720P.  
XX  
XX 27-SEP-2000; 2000US-0235840P.  
XX  
XX 27-SEP-2000; 2000US-0235863P.  
XX  
XX 28-SEP-2000; 2000US-0236028P.  
XX  
XX 28-SEP-2000; 2000US-0236032P.  
XX  
XX 28-SEP-2000; 2000US-0236033P.  
XX  
XX 28-SEP-2000; 2000US-0236034P.  
XX  
XX 28-SEP-2000; 2000US-0236109P.  
XX  
XX 28-SEP-2000; 2000US-0236111P.  
XX  
XX 29-SEP-2000; 2000US-0236842P.  
XX  
XX 29-SEP-2000; 2000US-0236891P.  
XX  
XX 02-OCT-2000; 2000US-0237172P.  
XX  
XX 02-OCT-2000; 2000US-0237173P.  
XX  
XX 02-OCT-2000; 2000US-0237288P.  
XX  
XX 02-OCT-2000; 2000US-0237294P.  
XX  
XX 02-OCT-2000; 2000US-0237295P.  
XX  
XX 02-OCT-2000; 2000US-0237316P.  
XX  
XX 03-OCT-2000; 2000US-0237425P.  
XX  
XX 03-OCT-2000; 2000US-0237598P.  
XX  
XX 03-OCT-2000; 2000US-0237604P.  
XX  
XX 03-OCT-2000; 2000US-0237606P.  
XX  
XX 03-OCT-2000; 2000US-0237608P.  
XX  
XX 01-NOV-2000; 2000US-0244867P.  
XX  
XX 01-NOV-2000; 2000US-0245084P.  
XX  
XX (AVAL-) AVALON PHARM.  
XX  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a chemical

PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX  
XX Claim 1; SEQ ID NO 4272; 4pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytosolic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
CC tumour  
XX  
XX Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
SQ  
Alignment Scores:  
Pred. No.: 2,06e-52 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 6 Gaps: 0  
US-09-701-674A-23 (1-179) x ABL65935 (1-697)  
Qy 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnrleuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGGTCAGTACATCGCCACTTCAGGCAGATCTCTGGCTAACGAC 62  
Qy 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValalagly 158  
Db 63 AAATACGAGAACGGGTACATTCCACCGGTCAACCTGACGTGGCCCTTTATGTGGCCGG 122  
Qy 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysglyThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGTGACCGGAGCGCTTATGTGACCACCGCG 182  
Qy 179 Ser 179  
Db 183 TCC 185  
RESULT 5  
ABK64784  
ID ABK64784 standard; DNA; 697 BP.  
XX  
XX ABK64784;  
AC  
XX  
XX 18-JUN-2002 (first entry)  
DT  
XX Human benign prostatic hyperplasia gene #679.  
DE  
XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
KW  
XX Homo sapiens.  
OS  
XX WO200212440-A2.  
XX  
XX 14-FEB-2002.  
PD  
XX  
XX 07-AUG-2001; 2001WO-US024708.  
XX  
XX 07-AUG-2000; 2000US-0223323P.  
XX  
XX 05-JUN-2001; 2001US-00873319.  
PR

XX (GENE-) GENE LOGIC INC.  
 PA (NLSB ) JAPAN TOBACCO INC.  
 XX Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
 XX WPI; 2002-257476/30.  
 XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 PT cells.  
 XX Disclosure; Page 367; 443pp; English.  
 XX The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK4106-ABK4860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention  
 XX  
 SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
 Alignment Scores:  
 Pred. No.: 2,066-52 Length: 697  
 Score: 61.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.08% Indels: 0  
 DB: 6 Gaps:  
 US-09-701-674A-23 (1-179) x ABK64784 (1-697)  
 QY 119 AspThrLeuArgLeuAlaSerSerrTrilleAlaHisLeuArgGlnIleLeuAlaSer 138  
 Db 3 GACACGCTCAGGTGGCGTCCAGTACATCGCCCACTTGAGGACAGATCTGGCTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValanLeuThrTrpProPheMetValaLagly 158  
 Db 63 AAATACGAGAACGGGTACATTACCCGGTCAACCTGACGTGGCCCTTTATGTGTGCCGGG 122  
 QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db 123 AAACCCGAGAGTGACCTGAAGAAGTGTGTGACCGGAGCGCGCTTATGTGGACCAACCGCG 182  
 QY 179 Ser 179  
 Db 183 TCC 185  
 RESULT 6  
 ID ABQ55427  
 AC ABQ55427 standard; cDNA; 632 BP.  
 XX  
 XX 22-AUG-2002 (first entry)  
 DT

XX Human ovarian antigen HNOJ185 cDNA, SEQ ID NO:1307.  
 DE Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 XX ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy; dysmenorrhea; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX Homo sapiens.  
 OS WO200200677-A1.  
 XX 03-JAN-2002.  
 XX 07-JUN-2001; 2001WO-US018569.  
 XX 07-JUN-2000; 2000US-0209467P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;  
 DR WPI; 2002-147878/19.  
 DR P-PSDB; ABP42350.  
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.  
 XX Claim 1; SEQ ID NO 1307; 2922pp; English.  
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system,  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 632 BP; 145 A; 176 C; 201 G; 93 T; 0 U; 17 Other;  
 Alignment Scores:  
 Pred. No.: 3,18e-40 Length: 632  
 Score: 49.00 Matches: 49

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 27.37% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABQ55427 (1-632)

QY 24 LysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluLysSerSerAsn 43  
 DB 82 AAATGGATTCGACCAAGGAATTGTGCTTCCACGAGACCGAGGAGCTCCAAAC 141  
 QY 44 CysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgGlyAlaPro 63  
 DB 142 TCGAGATGGTCTCTCCACAGAGGCGCGGCGCTTCCGCAAGAGGAGGAGCGGCC 201  
 QY 64 ThrLysLysSerProLeuSerGlyVal 72  
 DB 202 ACCAAGAGAGCCCTGAGCGGTGTC 228

RESULT 7  
 ACA56430  
 ID ACA56430 standard; cDNA; 240 BP.  
 AC ACA56430;  
 XX  
 XX 06-JUN-2003 (first entry)  
 DT  
 DE Chicken signalling pathway polynucleotide probe SEQ ID NO 1028.  
 XX  
 XX Chicken; probe; ss; array element; Parkinson's disease;  
 KW signalling pathway population; cancer; adenocarcinoma; leukaemia;  
 KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.  
 XX  
 OS Gallus gallus.  
 XX  
 XX US6500938-B1.  
 FN  
 PD 31-DEC-2002.  
 XX  
 XX 30-JAN-1998; 98US-00016434.  
 PF  
 PR 30-JAN-1998; 98US-00016434.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Au-Young J, Seilhamer JJ;  
 PI  
 XX WPI; 2003-352189/33.  
 DR  
 XX Combination of polynucleotide probes, useful as array elements in a  
 PT microarray for monitoring the expression of a number of target  
 PT polynucleotides.  
 XX  
 PS Claim 1; SEQ ID NO 1028; 65pp; English.  
 XX

*102(c) and odp.*

The invention relates to a combination which, comprises a number of polynucleotide probes comprising a sequence selected from one of the 1490 sequences mentioned in the specification. The combination is useful as an array element in a microarray for monitoring the expression of a number of target polynucleotides. The microarray is particularly useful in the diagnosis and treatment of cancer and immunopathology and neuropathology. The microarray is useful in diagnostics and treatment regimens, drug discovery and development, toxicological and carcinogenicity studies, forensics and pharmacogenomics. The microarray is also useful for monitoring progression of diseases and for developing sophisticated profiles for the effects of currently available therapeutic drugs. The combination is also useful for purifying a subpopulation of mRNAs, cDNAs and genomic fragments and in research and diagnostic applications. The array can detect changes in expression in a large number of genes coding for different signalling pathway populations which can be used to diagnose various diseases including cancer e.g. adenocarcinoma and leukaemia, CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease and Parkinson's disease. The present sequence represents a polynucleotide

CC probe of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format directly from USPIO at  
 CC segdata.uspto.gov/sequence.html?DocID=06500938B1  
 XX  
 SQ Sequence 240 BP; 55 A; 83 C; 72 G; 27 T; 0 U; 3 Other;

Alignment Scores:  
 Pred. No.: 2,24e-27 Length: 240  
 Score: 36.00 Matches: 36  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 20.11% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x ACA56430 (1-240)

QY 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTrpVal 109  
 DB 90 GCCCGCATGCGAGTGTGAGCAGAGGCTTCTCCAGACTCAAGACCCCTGCGCTGGGTG 149  
 QY 110 ProProAspThrLysLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 DB 150 CCCCCGACACCAAGCTCTCCAGCTGAGACGCTCAGGCTGGCTGGCTGCC 197

RESULT 8  
 ADI56226  
 ID ADI56226 standard; DNA; 240 BP.  
 XX  
 AC ADI56226;  
 XX  
 XX 22-APR-2004 (first entry)  
 DT  
 DE Human polynucleotide probe #1028.  
 XX  
 XX Human; probe; ss; receptor-like polypeptide; transducing polypeptide;  
 KW effector-like polypeptide; cancer; immunopathology; neuropathology;  
 KW drug development; toxicology; carcinogenicity;  
 KW signalling pathway polypeptide; adrenal gland; bladder; bone;  
 KW bone marrow; brain; breast; cervix; tumour; immunopathology; AIDS;  
 KW diabetes; pancreatitis; osteoporosis; ulcerative colitis; neuropathology;  
 KW dementia; amnesia; epilepsy; Alzheimer's disease; depression.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2004010136-A1.  
 FN  
 XX 15-JAN-2004.  
 PD  
 XX 26-NOV-2002; 2002US-00305720.  
 PF  
 XX 30-JAN-1998; 98US-00016434.  
 PR  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 XX Au-Young J, Seilhamer JJ;  
 PI  
 XX WPI; 2004-090520/09.  
 DR  
 XX New composition comprising polynucleotide probes, useful as array  
 PT elements in a microarray for monitoring the expression of target  
 PT polynucleotides or purifying a subpopulation of mRNAs, cDNA, or genomic  
 PT fragments.  
 XX  
 PS Claim 6; SEQ ID NO 1028; 73pp; English.  
 XX

The invention relates to a composition of polynucleotide probes comprising first polynucleotide probes comprising at least a portion of a gene encoding a receptor-like polypeptide, second polynucleotide probes comprising at least a portion of a gene encoding a transducing polypeptide and third polynucleotide probes comprising at least a portion of a gene encoding an effector-like polypeptide. The probes of the composition are useful as array elements in a microarray for monitoring



CC the expression of target polynucleotides. The microarray is useful in the  
CC diagnosis and treatment of cancer, an immunopathology or a  
CC neuropathology. It can also be used for drug discovery and development,  
CC toxicological and carcinogenicity studies, forensics or pharmacogenomics.  
CC Microarrays can also be used for monitoring the progression of diseases  
CC that may be associated with the altered expression of signalling pathway  
CC polypeptides. The composition can also be used to purify a subpopulation  
CC of mRNAs, cDNAs, or genomic fragments in a sample. The expression profile  
CC is also useful for the diagnosis and treatment of cancer, e.g. cancers of  
CC the adrenal gland, bladder, bone, bone marrow, brain, breast or cervix,  
CC an immunopathology, e.g. AIDS, diabetes, pancreatitis, osteoporosis or  
CC ulcerative colitis, or a neuropathology, e.g. dementia, amnesia,  
CC epilepsy, Alzheimer's disease or depression. This sequence represents a  
CC human polynucleotide probe of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX

SQ Sequence 240 BP; 55 A; 83 C; 72 G; 27 T; 0 U; 3 Other;

Alignment Scores:  
Pred. No.: 2,24e-27 Length: 240  
Score: 36.00 Matches: 36  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 20.11% Indels: 0  
DB: 12 Gaps: 0

US-09-701-674A-23 (1-179) x ADI56226 (1-240)

Qy 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTyrVal 109  
Db 90 GCCCGCATCGAGTGTGAGCAAGCCCTTCTCCAGACTCAAGACCACCTGCGCTGGGTG 149

Qy 110 ProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
Db 150 CCCCCGACACCAAGCTCTCCAGCTGGACACGCTCAGCTGGCTGCC 197

RESULT 9  
ACH23969  
ID ACH23969 standard; cDNA; 493 BP.  
XX  
AC ACH23969;  
XX

13-OCT-2003 (first entry)

Human adult ovary cDNA #2349.

Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
genome mapping; biodiversity; genetic disorder.  
XX  
OS Homo sapiens.  
XX

US2003073623-A1.

17-APR-2003.

30-JUL-2001; 2001US-00918995.

30-JUL-2001; 2001US-00918995.

(DRMA/) DRMANAC R T.

(LABA/) LABAT I.

(STAC/) STACHE-CRAIN B.

(DICK/) DICKSON M C.

(JONE/) JONES L W.

Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

WPI; 2003-615964/58.

New polynucleotide sequences obtained from various cDNA libraries, useful  
as hybridization probes, as oligomers for PCR, for chromosome and gene

PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.

PS Claim 1; SEQ ID NO 11181; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX

SQ Sequence 493 BP; 121 A; 136 C; 121 G; 112 T; 0 U; 3 Other;

Alignment Scores:  
Pred. No.: 6.28e-19 Length: 493  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x ACH23969 (1-493)

Qy 152 TrpProPheMetValAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSer 171

Db 102 TGGCCCTTTAATGTCGCGGAAACCCGAGAGTGACCTGAAGAGTGTGACCGGAGC 161

Qy 172 ArgLeuCysGlyThrThrAlaSer 179

Db 162 CGCTTATGTGAACACACCGCGTCC 185

RESULT 10

AAZ50465

ID AAZ50465 standard; cDNA; 1716 BP.

XX AAZ50465;

23-MAY-2000 (first entry)

Human muscudin bHLH protein encoding cDNA.

Muscudin; basic helix-loop-helix; bHLH; transcription factor; myogenesis;  
screening; myogenic disease; aberration; muscle development; cytostatic;  
gene therapy; diagnostic agent; muscular dystrophy; myopathy;  
neuromuscular; skeletomuscular; myogenic cancer; human; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 2..607

/\*tag= a

/product= "Human muscudin bHLH protein"

/note= "Lacks Alanine-Threonine myogenic recognition motif"

WO200006720-A1.

10-FEB-2000.

30-JUL-1999; 99WO-AU000623.



XX WPI; 2003-875398/81.  
 DR P-PSDB; ADE25769.  
 XX  
 PT Combination containing several polynucleotide that are differentially  
 PT expressed in foam cells and complements of the polynucleotides, useful  
 PT for diagnosing cardiovascular disease or atherosclerosis.  
 XX  
 PS Claim 1; SEQ ID NO 97; 37pp; English.  
 XX  
 CC The invention relates to a combination comprising several polynucleotides  
 CC having any one of 127 sequences (S1) such as the sequence of human  
 CC calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A4  
 CC hydrolase, human CGI-142 protein mRNA, human K+ channel beat 2 subunit  
 CC mRNA, etc., and their complements. The cDNAs are differentially expressed  
 CC in LPS (lipopolysaccharide)-treated foam cells. Also included are  
 CC obtaining an extended or full length gene from a library of nucleic acid  
 CC sequences, an expression vector containing the nucleic acids, a host cell  
 CC containing the vector, a purified polypeptide appearing as ADE25750 and  
 CC ADE25751, producing a protein by culturing the host cell, and a  
 CC composition comprising a purified antibody that specifically binds to the  
 CC proteins. The foam cell-expressed nucleic acids are useful for a high  
 CC throughput detection of differential expression of one or more  
 CC polynucleotides in a sample. The sample is from a subject with  
 CC atherosclerosis and comparison with a standard defines early, mid or late  
 CC stages of the disorder. The foam cell-expressed nucleic acids are useful  
 CC for high throughput screening of a library of molecules or compounds to  
 CC identify a ligand which binds a polynucleotide. The library is chosen  
 CC from DNA molecules, peptides, proteins and RNA molecules. The protein is  
 CC useful for a high throughput screening of library of molecules or  
 CC compounds to identify at least one ligand which specifically binds a  
 CC protein, for purifying a ligand from a sample for making an antibody. The  
 CC foam cell-expressed nucleic acids are useful for diagnosing  
 CC cardiovascular disorder. The foam cell-expressed nucleic acids are useful  
 CC as elements on a microarray which can be used for detecting related  
 CC polynucleotide in a sample, diagnosing cardiovascular disease,  
 CC atherosclerosis. The present sequence represents a cDNA whose expression  
 CC is upregulated in LPS treated foam cells.  
 XX  
 SQ Sequence 2177 BP; 423 A; 691 C; 650 G; 413 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,576-18 Length: 2177  
 Score: 28.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.64% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-701-674A-23 (1-179) x ADE25693 (1-2177)  
 QY 106 LeuProTTPValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 DB 857 CTGCCCTGGTGCCCGCCGACACTAAGCTCTCCAGCTGGACGCTCCGGCTGCTTCC 916  
 QY 126 SerTyrlleAlaHisLeuArgGln 133  
 DB 917 AGTTACATCGCTCACCTGGCGAG 940  
 RESULT 13  
 AAA72428  
 ID AAA72428 standard; cDNA; 2196 BP.  
 XX  
 AC AAA72428;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Human nucleic acid-binding protein NuABP-47 cDNA.  
 XX  
 KW Human nucleic acid-binding protein; NuABP; agonist; antagonist; EST;  
 KW expressed sequence tag; drug screening; recombinant expression; antibody;  
 KW reproductive disorder; infertility; immunological disorder;  
 KW neurological disorder; cell proliferative disorder; cancer; tumour; ss.

XX Homo sapiens.  
 OS  
 PN WO200044900-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 28-JAN-2000; 2000WO-US002237.  
 XX  
 PR 29-JAN-1999; 99US-0117904P.  
 PR 29-JAN-1999; 99US-0117905P.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Tang YT, Lal P, Hillman JL, Yue H, Azimzai Y, Lu AMD, Baughn MR;  
 PI Tran B, Shih LL, Au-Young JL;  
 XX  
 DR WPI: 2000-499332/44.  
 DR P-PSDB; AAB21043.  
 XX  
 PT Novel nucleic acid binding proteins, used to identify agonists and  
 PT antagonists of them, for the treatment of reproductive, immunological,  
 PT neurological and cell proliferative disorders including cancer.  
 XX  
 PS Claim 4; Page 176; 180pp; English.  
 XX  
 CC Sequences AAA72382-A72436 represent cDNAs encoding novel human nucleic  
 CC acid-binding proteins (NuABPs; AAB20997-B21051). These cDNAs were  
 CC produced by extension from an appropriate EST (expressed sequence tag)  
 CC using primers designed using the EST. The invention also relates to  
 CC expression constructs, host cells and transgenic organisms comprising a  
 CC human NuABP nucleic acid, recombinant production of the human NuABPs, and  
 CC antibodies against the human NuABPs, and also to methods of screening  
 CC modulators of human NuABP activity or expression. The human NuABPs, and  
 CC their agonists and antagonists are used to treat diseases associated with  
 CC overexpression or underexpression of functional NuABPs. Human NuABP  
 CC proteins and nucleotides, and NuABP agonists and antagonists can be used  
 CC to diagnose, treat and prevent reproductive, immunological, neurological  
 CC and cell proliferative disorders. Reproductive disorders that may be  
 CC treated using compositions of the invention include infertility,  
 CC endometriosis, disruptions of the menstrual cycle and disruptions of  
 CC spermatogenesis. Immunological disorders that may be treated include  
 CC AIDS, allergies, and autoimmune disorders such as multiple sclerosis,  
 CC rheumatoid arthritis, diabetes and systemic lupus erythematosus.  
 CC Neurological disorders that may be treated include epilepsy,  
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's  
 CC disease, prion diseases such as Creutzfeldt-Jakob disease, and mental  
 CC disorders such as schizophrenia. Cell proliferative disorders that may be  
 CC treated include a wide variety of cancers, and also arteriosclerosis,  
 CC atherosclerosis, cirrhosis and psoriasis  
 XX  
 SQ Sequence 2196 BP; 438 A; 691 C; 653 G; 414 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,596-18 Length: 2196  
 Score: 28.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.64% Indels: 0  
 DB: 3 Gaps: 0  
 US-09-701-674A-23 (1-179) x AAA72428 (1-2196)  
 QY 106 LeuProTTPValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 DB 857 CTGCCCTGGTGCCCGCCGACACTAAGCTCTCCAGCTGGACGCTCCGGCTGCTTCC 916  
 QY 126 SerTyrlleAlaHisLeuArgGln 133  
 DB 917 AGTTACATCGCTCACCTGGCGAG 940  
 RESULT 14  
 ADJ56314

```

ID ADJ56314 standard; cDNA; 2196 BP.
XX AC ADJ56314;
XX DT 06-MAY-2004 (first entry)
XX DE Human cDNA differentially expressed in MYCN activated cells SeqID 120.
XX KW human; differential expression; transactivator; proto-oncogene;
XX KW neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;
XX KW MYCN activated cell.
XX OS Homo sapiens.
XX PN US2003119009-A1.
XX PD 26-JUN-2003.
XX PF 25-FEB-2002; 2002US-00084817.
XX PR 23-FEB-2001; 2001US-0270784P.
XX PA (STUN/) STUART S G.
XX PA (NUCH/) NUCHTERN J G.
XX PA (PLON/) PLON S E.
XX PA (SHOH/) SHOHET J M.
XX PI Stuart SG, Nuchtern JG, Plon SE, Shohet JM;
XX WPI; 2003-635698/60.
XX PT New genes regulated by MYCN activation, useful in gene therapy,
XX PT particularly for treating a subject with e.g. neuroblastoma or other
XX PT cancers, or for diagnosing, staging or monitoring the treatment of the
XX PT cancer.
XX PS Claim 1; SEQ ID NO 120; 27pp; English.
XX CC This invention relates to novel isolated cDNAs that are differentially
XX CC expressed in MYCN activated cells. Specifically, it refers to
XX CC polynucleotide sequences that exhibit differential expression patterns in
XX CC cells activated by the transactivator MYCN, where MYCN is a proto-
XX CC oncogene that is amplified in neuroblastoma cells and is common in small
XX CC cell lung cancers. The present invention describes these cDNA molecules
XX CC as useful for in hybridisation assays to detect expression of nucleic
XX CC acids (or complementary nucleic acids) in a present in a given sample, as
XX CC well as for screening assays by identifying molecules or compounds that
XX CC specifically bind the cDNA as a ligand and modulate function or activity.
XX CC Accordingly, these compositions exhibit cytostatic activity and can also
XX CC be used for gene therapy purposes. This polynucleotide sequence is a cDNA
XX CC that is differentially expressed in MYCN activated cells, given in an
XX CC exemplification of the invention. NOTE: This sequence does not appear in
XX CC the printed specification but has been obtained in electronic format from
XX CC the US Patent Office at
XX CC ftp.segdata.uspto.gov/sequence.html?DocID=20030119009.
XX SQ Sequence 2196 BP; 438 A; 691 C; 653 G; 414 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,598-18 Length: 2196
Score: 28.00 Matches: 28
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.64% Indels: 0
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x ADJ56314 (1-2196)
QY 106 LeuProTtpValProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
DB 857 CTGCCCTGGTCCCGCCGACACTAAGCTCTCCAGCTGGACAGCTCCGGCTGCTTCC 916
QY 126 SerTyrIleAlaHisLeuArgGln 133

US-09-701-674A-23 (1-179) x AAS94828 (1-2382)
QY 106 LeuProTtpValProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
DB 861 CTGCCCTGGTCCCGCCGACACTAAGCTCTCCAGCTGGACAGCTCCGGCTGCTTCC 920
QY 126 SerTyrIleAlaHisLeuArgGln 133
DB 921 AGTTACATCGCTCACCTGGCGCAG 944

```

```

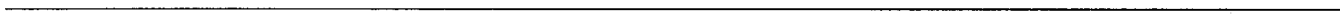
Db 917 AGTTACATCGCTCACCTGGCGCAG 940
RESULT 15
AAS94828
ID AAS94828 standard; DNA; 2382 BP.
XX AC AAS94828;
XX DT 14-FEB-2002 (first entry)
XX DE Human DNA sequence #83 expressed during foam cell differentiation.
XX KW Human; foam cell differentiation; atherosclerosis; cerebral stroke;
XX KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX OS Homo sapiens.
XX PN WO200177389-A2.
XX PD 18-OCT-2001.
XX PF 04-APR-2001; 2001WO-US011128.
XX PR 05-APR-2000; 2000US-0195106P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
XX PI Tai J;
XX WPI; 2002-010925/01.
XX PT Composition useful for diagnosis of conditions, disorders or diseases
XX PT associated with atherosclerosis, comprises several polynucleotides that
XX PT are differentially expressed in foam cell development.
XX PS Claim 1; Page 135-136; 315pp; English.
XX CC The present invention relates to the isolation of human polynucleotide
XX CC sequences that are differentially expressed during foam cell
XX CC differentiation. The polynucleotide sequences of the invention or a
XX CC composition comprising these polynucleotides are useful as a high
XX CC throughput method for detecting altered expression of one or more
XX CC polynucleotides in a sample. The polynucleotides can be used in the
XX CC diagnosis of disorders associated with foam cell development such as
XX CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
XX CC coronary artery disease. The polynucleotide sequences can also be used as
XX CC PCR primers and probes. The polynucleotides of the invention are also
XX CC useful in gene therapy. AAS94746-AAS95021 represent the human
XX CC polynucleotide sequences of the invention which are differentially
XX CC expressed during foam cell differentiation.
XX SQ Sequence 2382 BP; 459 A; 732 C; 702 G; 489 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,8e-18 Length: 2382
Score: 28.00 Matches: 28
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.64% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x AAS94828 (1-2382)
QY 106 LeuProTtpValProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
DB 861 CTGCCCTGGTCCCGCCGACACTAAGCTCTCCAGCTGGACAGCTCCGGCTGCTTCC 920
QY 126 SerTyrIleAlaHisLeuArgGln 133
DB 921 AGTTACATCGCTCACCTGGCGCAG 944

```

Search completed: December 17, 2004, 01:22:05  
Job time : 461 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 19:03:27 ; Search time 728 Seconds

(without alignments)  
9646.299 Million cell updates/sec

Title: US-09-701-674A-54

Perfect score: 1272

Sequence: 1 tctacggccagcactctggg.....attgctataaaaaaaaaa 1272

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4093002 seqs, 2760418825 residues

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581	45.7	697	9	US-09-954-456-829
2	581	45.7	697	9	Sequence 829, App
3	581	45.7	697	10	Sequence 1245, App
4	581	45.7	697	10	Sequence 1028, App
5	525.6	41.3	632	16	Sequence 679, App
6	466.2	36.7	471	9	US-10-264-049-1307
C	466.2	36.7	471	9	Sequence 1307, App
C	466.2	36.7	471	9	Sequence 881, App
C	466.2	36.7	471	9	Sequence 1431, App
C	466.2	36.7	471	10	US-09-954-456-1431
C	466.2	36.7	471	10	Sequence 17, App
9	466.2	36.7	471	10	US-09-873-319-10
10	434.2	34.1	493	10	US-09-918-995-11181
11	297.2	23.4	482	10	US-09-918-995-9069
12	213	16.7	240	16	US-10-305-720-1028

13	211.6	16.6	2177	15	US-10-247-671-97	Sequence 97, Appl
14	211.6	16.6	2196	15	US-10-084-817-120	Sequence 120, App
15	210	16.5	1716	17	US-10-775-169-201	Sequence 201, App
16	208.6	16.4	2382	15	US-10-240-965-83	Sequence 83, Appl
17	198	15.6	1939	17	US-10-755-889-365	Sequence 365, App
18	81.8	6.4	446	9	US-09-833-381-505	Sequence 505, App
19	78.4	6.2	1140	17	US-10-450-826-124	Sequence 124, App
C	71.8	5.6	609	15	US-10-029-386-22740	Sequence 22740, A
20	71.8	5.6	1800	10	US-09-771-357-106	Sequence 106, App
21	71.8	5.6	1800	15	US-10-059-579-106	Sequence 106, App
22	71.8	5.6	1800	15	US-10-059-579-106	Sequence 106, App
23	71	5.6	450	13	US-10-004-717-22	Sequence 22, Appl
24	71	5.6	450	18	US-10-860-373-22	Sequence 22, Appl
25	71	5.6	450	18	US-10-860-373-22	Sequence 22, Appl
26	69.4	5.5	1457	9	US-09-954-531-982	Sequence 982, Appl
27	69.4	5.5	1467	16	US-10-062-674-1950	Sequence 1950, App
28	69.4	5.5	1678	15	US-10-240-965-119	Sequence 119, App
C	69	5.4	158405	14	US-10-175-523-86	Sequence 86, Appl
30	68.8	5.4	2537	16	US-10-466-164-25	Sequence 25, Appl
31	68.8	5.4	2550	18	US-10-335-053-106	Sequence 106, App
32	68.8	5.4	2601	9	US-09-880-107-2433	Sequence 2433, App
33	68.2	5.4	82660	13	US-10-087-192-2017	Sequence 2017, App
C	68.2	5.4	83493	18	US-10-331-059-61	Sequence 61, Appl
35	68	5.3	651	9	US-09-749-728B-22	Sequence 22, Appl
36	67.6	5.3	1132	9	US-09-778-844-23	Sequence 23, Appl
C	67.6	5.3	256525	13	US-10-087-192-451	Sequence 451, App
38	67.4	5.3	444	18	US-10-674-124A-6324	Sequence 6324, App
39	67.4	5.3	2466	13	US-10-044-090-251	Sequence 251, App
C	66.8	5.3	486	10	US-09-918-995-22633	Sequence 22633, A
C	66.2	5.2	505	10	US-09-918-995-25117	Sequence 25117, A
42	66.2	5.2	42999	9	US-09-799-462A-17	Sequence 17, Appl
43	66.2	5.2	42999	10	US-09-836-911A-17	Sequence 17, Appl
44	66.2	5.2	42999	10	US-09-738-630-73	Sequence 73, Appl
45	66.2	5.2	42999	13	US-10-125-767-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-954-456-829  
; Sequence 829, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954, 456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233, 617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234, 052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234, 923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235, 637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235, 711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235, 720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235, 840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235, 863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 829  
; LENGTH: 697







```

Db 481 AGAGCGCTATCAATGTAATCTTTTGTACAAATATGTTGTAATAATGTAGATCATAGATAGCT 540
QY 1159 GACTTTGACAGTCACATTTATTAAGTAATTCACCTTAAGATATATATTTTCAACAA 1218
Db 541 GACTTTGACAGTCACATTTATTAAGTAATTCACCTTAAGATATATATTTTCAACAA--- 597
QY 1219 GTTTTGCTACTTTTGAATAAATCTTTCTTTATATATCTTAAATAAAAAA 1271
Db 598 -AGTTGCACTTTTGAATAAATCTTTCTTTATATCTTAAATAAAAAAAGA 649

```

## RESULT 5

```

US-10-264-049-1307
; Sequence 1307, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1307
; LENGTH: 632

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (78)..(79)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (174)..(174)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231)..(231)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (249)..(249)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (263)..(263)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (265)..(265)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (280)..(280)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (331)..(331)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (433)..(433)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (467)..(467)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (498)..(498)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (537)..(537)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (546)..(546)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (552)..(552)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (623)..(623)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (629)..(629)
; OTHER INFORMATION: n equals a,t,g, or c
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-1307

Query Match 41.3%; Score 525.6; DB 16; Length 632;
Best Local Similarity 93.4%; Pred. No. 1.6e-118;
Matches 593; Conservative 0; Mismatches 36; Indels 6; Gaps 5;

QY 257 CTCCTCCCAACATGTCACCGGCTCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGA 316
Db 2 CTCCTCCCAACATGTCACCGGCTCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGA 61
QY 317 TGTGGAATGTGACGGTTGAAAATGGAATCGAAACAAGAAATTTGTACTTCAACGAGA 376
Db 62 TGTGGAATGTGACGGNTGAAAATGGAATCGAAACAAGAAATTTGTACTTCAACGAGA 121
QY 377 GCACCGAGGAGCTCCAACTGGGAGATGGTCTCCCGAGAGGGCCGCGCGCCCTGG 436
Db 122 GCACCGAGGAGCTCCAACTGGGAGATGGTCTCCCGAGAGGGCCGCGCGCCCTGG 181
QY 437 GCAAGAGAGGAAGCGCGCCACCAAGAGAGAGCGCCCTCAGCGGGTCAAGAGAGGGA 496
Db 182 GCAAGAGAGGAAGCGCGCCACCAAGAGAGAGCGCCCTCAGCGGGTCAAGAGAGGGA 241
QY 497 AGCAGTCCAGGCAACGCGCGCAAGCGCGAGAGGGGCCCGCATGCGAGTCTGAGCA 556
Db 242 AGCAGTCCAGGCAACGCGCGCAAGCGCGAGAGGGGCCCGCATGCGAGTCTGAGCA 301
QY 557 AGGCTTTCTCCAGACTCAAGACACCCCTGCTGGTGCCCGCCCGACCAAGCTCTCA 616
Db 302 AGGCTTTCTCCAGACTCAAGACACCCCTGCTGGTGCCCGCCCGACCAAGCTCTCA 361
QY 617 AGCTGACACGCTCAGCTGGGTCAGCTACATGCCCACTTGAAGGAGATCTCTGGCTA 676
Db 362 AGCTGACACGCTCAGCTGGGTCAGCTACATGCCCACTTGAAGGAGATCTCTGGCTA 421
QY 677 AGCACAATACGAGACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTG 736
Db 422 ACACCAATACGAGACGGGTACATTCACCGGTCAACCTGACGTGGCCCTTTATGGTG 479
QY 737 CCGGGAACCCGAGAGTGACCTGAAAGAGAGTGG-TGACCGGAGCGCTTATGTGAACC 795
Db 480 -CGGGAACCCGAGAGTGACCTGAAAGAGAGTGGTGGACCGGAGCGGTTATGTGAANC 538
QY 796 ACCGCTCTGACCTTGGAGTGGAGTGTGGAAAGCGCGCTCCCGGGGAGCGGC 855
Db 539 ACCGCTCTGACCTTGGAGTGGAGTGTGGAAAGCGCGCTCCCGGGGAGCGGC 596
QY 856 CCGGGAAGCGGACCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 890
Db 597 CCGGGAAGCGGACCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 631

```





```

; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11181
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(493)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-11181

Query Match          34.1%;   Score 434.2;   DB 10;   Length 493;
Best Local Similarity 99.3%;   Pred No. 3.8e-96;
Matches 436;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0

QY      674  CTAACGACAAATACGAGAACCGGTACATTCACCCGGTCAACTGACGTGGCCCTTTATGG 733
Db      55   CTAACGACAAATACGAGAACCGGTACATTCACCCGGTCAACTGATGTGGCCCTTTATGG 114

QY      734  TGGCCGGGAACCCGAGAGTGCACCTGAAGAAGTGGTGACCGCGAGCGCTTATGTGAA 793
Db      115  TGGCCGGGAACCCGAGAGTGCACCTGAAGAAGTGGTGACCGCGAGCGCTTATGTGAA 174

QY      794  CCACCGGCTCTGCACCTTGGAGGTCCGAGTCTGGGAAAGGCGCGTCCCGGGGAGCGG 853
Db      175  CCACCGGCTCTGCACCTTGGAGGTCCGAGTCTGGGAAAGGCGCGTCCCGGGGAGCGG 234

QY      854  GCCCGGGAAGGAGACCCCTGCCCTCAGTGTCTCTGTCTGTCTGTCTTCCCTCGCAATGC 913
Db      235  GCCCGGGAAGGAGACCCCTGCCCTCAGTGTCTCTGTCTGTCTGTCTTCCCTCGCAATGC 294

QY      914  TCCTCTCTCTGTCCACCCCGGAGAACACTTTACAACGACGAGGAGATTCGTTTCCAAA 973
Db      295  TCCTCTCTCTGTCCACCCCGGAGAACACTTTACAACGACGAGGAGATTCGTTTCCAAA 354

QY      974  CCAGAGAGATCAATTGTACTTTACAAAGATTCCTCATTTTAACTTTTAACTTCTTAC 1033
Db      355  CCAGAGAGATCAATTGTACTTTACAAAGATTCCTCATTTTAACTTTTAACTTCTTAC 414

QY      1034  CGTCAATGACTCTCAAGCCTTGTGTGTCGAAGTGCATATGTAATTATAATATATAA 1093
Db      415  CGTCAATGACTCTCAAGCCTTGTGTGTCGAAGTGCATATGTAATTATAATATATAA 474

QY      1094  TAGATAAGAGCCTATCAAT 1112
Db      475  TAGATAAGAGCCTATCAAT 493

RESULT 11
US-09-918-995-9069
; Sequence 9069, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9069
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(482)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-9069

```

Qy 626 CGCTCAGGTGGCTCCAGTACATCGCCCACTTGAGGCAGATCTTGGCTAACGACAAAT 685

Db 181 CGCTCAGGTGGCTCCAGTACATCGCCCACTTGAGGCAGATCTTGGCTAACGACAAAT 240

```

RESULT 13
US-10-247-671-97
; Sequence 97, Application US/10247671
; Publication No. US20030194721a1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
; SEQ ID NO 97
; LENGTH: 2177
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20030194721a1 3282941CB1
US-10-247-671-97

```

Query Match	16.6%;	Score 211.6;	DB 15;	Length 2177;
Best Local Similarity	68.5%;	Pred. No. 3.6e-41;		
Matches 292;	Conservative 0;	Mismatches 134;	Indels 0;	Gaps 0
QY	377	GCACCGAGGAGAGCTCCAAC	TGCGAGAATGGGTCTCC	CACAGAGGGCGCGCGGCGCTGG 436
DB	651	GCAGCGCGAAGCTGCAAGAGGA	CGCGCCCGTGTGGGCGCGCGCAGGTG 710	
QY	437	GCNAGGAGGAAGGGGCCACCA	RAGAAGAGCCCTTGAGCGGGGT	CAGCAGAGGGGA 496
DB	711	GTAGCGCGGCGGTGGTGCAAG	AGGCCCTCCCGGCCAAGGGCT	CAGCGCAGAGTGCA 770
QY	497	AGCAGGTCCAGCGCAACGCGG	CCACGCGAGAGCGGCGCCGAT	CGCAGTGTCTGAGCA 556
DB	771	AGCAGTTCGAGCGGAACGCGG	CCACGCGCCAGCCGTGAGCGTG	CCCGATGCGGTCTGAGCA 830
QY	557	AGGCTTTCTCAGATCTAAGACC	ACCTTGCCCTGGGTGCCCCCGCA	CAACCAAGCTCTCA 616
DB	831	AAGCCTTCTCCAGAGTCAAGAC	CAAGCTGCGCTGGGTGCCCCCGCA	CACTAAGCTCTCCA 890
QY	617	AGCTGACACGCTCAGGCTGGG	GTGGGTCCAGCTACATCGCCACT	TTGAGGCAGATCTCGGCTA 676
DB	891	AGCTGGAACAGTCCGGCTGGG	TTCCAGTTACATGCTCACTT	CGCGCAGCTGTTCGAGG 950
QY	677	ACGACAATAACAGAACGGGTACA	TTACCCGGGTCAACTGACGTGG	CCCTTTATGGTGG 736
DB	951	AGGACCGCTATGAGAACGGCT	ACGTGTCACCGAGTAACCTGAC	ATGAGCCCAATTCGTGGTCT 101
QY	737	CGGGGAACCCGAGAGTGACCT	TGAAGAAGTGGTGACCGGAG	CGCGCTTATGTGGGAACCA 796
DB	1011	CGGGAAGACCGGACTCTGACCA	AAAGAGTTTCGGCAGCCACAG	ACTATGTGGGAACCA 1076
QY	797	CCGCGT	802	
DB	1071	CGGCTT	1076	

RESULT 14  
US-10-084-817-120  
; Sequence 120, Application US/10084817  
; Publication No. US20030119009A1  
; GENERAL INFORMATION:

```
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jasen M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 120
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 3282941CB1
US-10-084-817-120

Query Match          16.6%   Score 211.6; DB 15; Length 2196;
Best Local Similarity 68.5%; Pred.No. 3.6e-41;
Matches 292; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY      377 GCACCGAGGAGAGCTCCAATCGAGAAATGGGTCTCCCCAGAAGGGCCGCGCGCGCTGG 436
Db       651 GCAGCGGGAAGGTGCAAGAGGAAGCGGCCCGTGTGCCTGGGGCGCGCGCAGGTG 710

QY      437 GCAAGAGGAGNAGGCCGCCACCACAAGAGCCCCTCTGAGCGGGTCAGCCAGAGGGGA 496
Db       711 GTAGCGGGCGGTGTGGTGGCAAAGACCCCTCTCCGGGCCAAGGGCTCACGCCAGAGTGCA 770

QY      497 AGCAGGCTCCAGCGCAACGCGCGCAACGCGCGAGAGCGGCGCCGCAATGCAGTGTGTGACA 556
Db       771 AGCAGTCCGACGGGAACGCGGCGCAACGCGCGTGTGAGCGTGTCCCGATGCGTGTGAGCA 830

QY      557 AGGCGTTCTCCAGACTCAAGACCACCTTCGCTGGTGGTGGCCGCCCGACACCAAGCTCTCCA 616
Db       831 AAGCTTCTCCAGGCTCAAGACCAGGCTGCCCTGGTGGTGGCCGCCCGACACTAAGCTCTCCA 890

QY      617 AGCTGGACAGCGCTCAGGCTGGCGTCCAGGTATCATCGCCCACTTGAGGCAGATCCTGGCTA 676
Db       891 AGCTGGACAGCGCTCCGGCTGGCTTCAGTTATCATCGTCACTCGCGGCGAGCTGTTGCAGG 950

QY      677 AGCAATAATCGAAGACGGGTACATTTCACCCGGTCAACTGACGTGGCGCCTTAATGTTGG 736
Db       951 AGGACCGGTATGAAACCGGCTAGCTGCACCCAGTGAACCTGACATGGCCATTCGTGGTCT 1010

QY      737 CGGGAAACCCGAGGTGACTCAAAGAAGTGGTGACCCCGAGCGCGCTTATGTGGAACCA 796
Db     1011 CGGAGNACCGGACTCTGACACCAAGAAGTTTCCGCGAGCCACAGACTATGTGGAACCA 1070

QY      797 CCGCGT 802
Db     1071 CCGCTT 1076
```

RESULT 15  
US-10-775-169-201  
; Sequence 201, Application US/10775169  
; Publication No. US20040175743A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dörner, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 201

; LENGTH: 1716

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-775-169-201

Query Match 16.5%; Score 210; DB 17; Length 1716;

Best Local Similarity 68.3%; Pred. No. 7.8e-41;

Matches 291; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY	377	GCACCGAGGAGAGCTCCAACTGGGAGATGGGTCTCCCCAGAAAGGSCCGCGGCTGG	436
Db	180	GCAGCCGGAAGGCTGCAAGAGGAAGCGGCCCGCGGSGCTGGGGCGCGCGCAGGTG	239
QY	437	GCAAGAGGAGGAAGGCGCCCAACAAAGAGGCCCCCTGAGCGGGTCAGCCAGGAGGGGA	496
Db	240	GTAGCGCGGCGGTGGTGCAAGAGCCCTCCCGGCCAAGGGCTCAGCGCAGAGTGCA	299
QY	497	AGCAGCTCCAGCGCAACGCGCGGAGAGCGGCGCCGATGCGAGTGCTGAGCA	556
Db	300	AGCAGTCGAGCGGAACGCGGCCACGCCCGTAGCGTGCCCGGATGCGCGTGTGAGCA	359
QY	557	AGGCCTTCTCCAGACTCAAGACCAACCTGCTGCTGGTGCCCCCGACACCAAGCTCTCCA	616
Db	360	AAGCCTTCTCCAGGCTCAAGACCAGCCTGCCCTGGGTGCCCGCGACCTAAGCTCTCCA	419
QY	617	AGCTGACACGCTCAGGCTGGGCTCCAGCTACATGCCCCACTGAGGCGAGATCCTGGCTA	676
Db	420	AGCTGACACGCTCCGGCTGGGCTTCCAGTTACATCGCTCACCTGCGGCGAGTGTTCAGG	479
QY	677	ACGCAAAATACGAGAACGGGTACATTACCCCGTCAACCTGACGTGGCCCTTTATGGTGG	736
Db	480	AGGACCGCTATGGAACGCTAGCTGCACCCAGTGNACTGACATGGCCATTCGTGGTCT	539
QY	737	CCGGGAACCCGAGAGTACCTGAAAGAGTGGTGACCGCGAGCCGCTTATGTGAACCA	796
Db	540	CGGGAGACCGGACTCTGACACCAAGAAGTTTCCGACGCCAACACAGACTATGTGGAACCA	599
QY	797	CGGCGT 802	
Db	600	CGGCTT 605	

Search completed: December 16, 2004, 22:04:49

Job time : 731 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 01:07:28 ; Search time 88 Seconds  
(without alignments)  
1445.809 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179

Sequence: 1 MSTGSLSDVEDLQEVEMLEC.....PESDLKEVTVASRLGGTTAS 179

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO spooll\_h/US09701674/runat\_15122004\_100531\_28340/app query.fasta\_1.327  
-DB=Issued Patents NA -QEXT=fastap -SUFFIX=oligo.xmi -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 @CGN 1 1 69 @runat\_15122004\_100531\_28340 -NCPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DMM TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*\*

1: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PTUS COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	20.1	240	US-09-016-434-1028	Sequence 1028, Ap
2	9	5.0	2610	US-09-489-039A-1390	Sequence 1390, Ap
3	8	4.5	276	US-09-270-767-3094	Sequence 3094, Ap
4	8	4.5	276	US-09-270-767-18376	Sequence 18376, A
5	8	4.5	633	US-09-328-111-493	Sequence 493, App
6	8	4.5	1013	US-09-270-767-10281	Sequence 10281, A
7	8	4.5	1893	US-09-489-039A-5852	Sequence 5852, Ap
8	8	4.5	2595	US-09-489-039A-5698	Sequence 5698, Ap
9	8	4.5	4853	US-08-793-824-1	Sequence 1, Appli
10	7	3.9	51	US-09-443-199C-469	Sequence 469, App
11	7	3.9	88	US-09-237-712-91	Sequence 91, Appl
12	7	3.9	151	US-09-513-999C-14487	Sequence 14487, A

13	7	3.9	158	4	US-09-513-999C-20456	Sequence 20456, A
c 14	7	3.9	176	4	US-09-313-294A-5167	Sequence 5167, Ap
15	7	3.9	216	4	US-09-702-705-735	Sequence 735, App
16	7	3.9	216	4	US-09-736-457-735	Sequence 735, App
17	7	3.9	216	4	US-09-614-124B-735	Sequence 735, App
18	7	3.9	216	4	US-09-671-325-735	Sequence 735, App
19	7	3.9	216	4	US-09-589-184-735	Sequence 735, App
20	7	3.9	216	4	US-09-658-824-735	Sequence 735, App
c 21	7	3.9	252	4	US-09-248-786A-13618	Sequence 13618, A
22	7	3.9	270	4	US-09-252-991A-13449	Sequence 13449, A
c 23	7	3.9	283	4	US-09-313-294A-6600	Sequence 6600, Ap
24	7	3.9	285	4	US-09-252-991A-9893	Sequence 9893, App
c 25	7	3.9	285	4	US-09-252-991A-16285	Sequence 16285, A
26	7	3.9	324	4	US-09-621-976-18476	Sequence 18476, A
c 27	7	3.9	367	3	US-09-085-199B-38	Sequence 38, Appl
28	7	3.9	375	4	US-09-513-999C-1244	Sequence 1244, Ap
29	7	3.9	380	4	US-09-513-999C-13010	Sequence 13010, A
30	7	3.9	399	4	US-09-489-039A-2211	Sequence 2211, Ap
31	7	3.9	405	4	US-09-621-976-14307	Sequence 14307, A
32	7	3.9	412	4	US-09-621-976-12072	Sequence 12072, A
33	7	3.9	455	4	US-09-513-999C-11196	Sequence 11196, A
34	7	3.9	461	4	US-09-270-767-5677	Sequence 5677, Ap
35	7	3.9	461	4	US-09-270-767-20959	Sequence 20959, A
c 36	7	3.9	465	4	US-09-252-991A-15919	Sequence 15919, A
c 37	7	3.9	477	4	US-09-489-039A-1796	Sequence 1796, Ap
c 38	7	3.9	477	4	US-09-270-767-7639	Sequence 7639, Ap
c 39	7	3.9	477	4	US-09-270-767-22921	Sequence 22921, A
c 40	7	3.9	486	3	US-09-328-111-443	Sequence 443, App
c 41	7	3.9	493	4	US-09-702-705-193	Sequence 193, App
c 42	7	3.9	493	4	US-09-736-457-193	Sequence 193, App
c 43	7	3.9	493	4	US-09-614-124B-193	Sequence 193, App
c 44	7	3.9	493	4	US-09-671-325-193	Sequence 193, App
c 45	7	3.9	493	4	US-09-589-184-193	Sequence 193, App

#### ALIGNMENTS

RESULT 1  
US-09-016-434-1028  
; Sequence 1028, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555

```

; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1028:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 240 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   LIBRARY: KIDNOT05
;   CLONE: 954226
; IMMEDIATE SOURCE:
; ORGANISM: Drosophila melanogaster
US-09-016-434-1028

Alignment Scores:
Pred. No.: 4,74e-29 Length: 240
Score: 36.00 Matches: 36
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.11% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-016-434-1028 (1-240)

Qy 90 AlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrLeuProTrpVal 109
Db 90 GCCCGCATCGAGTGTGAGCAAGGCTTCTCCAGACTCAAGACCACCTGCCCTGGGTG 149

Qy 110 ProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125
Db 150 CCCCCGACCAAGCTCTCCAAGTGGACACGCTCAGCGTGGCGTCC 197

RESULT 2
US-09-489-039A-1390/c
; Sequence 1390, Application US/09489039A
; Patent No. 6810836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1390
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1390

Alignment Scores:
Pred. No.: 10 Length: 2610
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.03% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-489-039A-1390 (1-2610)

Qy 52 GlyArgGlyLeuGlyLysArgArg 60
Db 1239 GGTCGGGTGGCTTGGTAAACGGCGT 1213

RESULT 3
US-09-270-767-3094/c
; Sequence 3094, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18376
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-18376/c
; Sequence 18376, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18376
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-18376

Alignment Scores:
Pred. No.: 12.9 Length: 276
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-270-767-18376 (1-276)

Qy 87 ArgGluArgAlaArgMetArgVal 94
Db 261 AGAGAGAGACGAAGATGCGCGTG 238

RESULT 4
US-09-270-767-18376/c
; Sequence 18376, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18376
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-18376

Alignment Scores:
Pred. No.: 12.9 Length: 276
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-270-767-18376 (1-276)

Qy 87 ArgGluArgAlaArgMetArgVal 94
Db 261 AGAGAGAGACGAAGATGCGCGTG 238

RESULT 5
US-09-328-111-493/c
; Sequence 493, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

```

```

; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; PRIOR FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 493
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(633)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-493

```

```

Alignment Scores:
Pred. No.: 28.7 Length: 633
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 3 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-328-111-493 (1-633)

```

QY 105 ThrLeuProTrrpValProasp 112
DB 470 ACTTACCATGGGTTCCCCACAG 447

```

#### RESULT 6

```

US-09-270-767-10281/c
; Sequence 10281, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10281
; LENGTH: 1013
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-10281

```

```

Alignment Scores:
Pred. No.: 45.2 Length: 1013
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-270-767-10281 (1-1013)

```

QY 119 AspThrLeuArgLeuAlaSerSer 126
DB 706 GACACTCTGAGGCTTGCCTCCCTCC 683

```

#### RESULT 7

```

US-09-489-039A-5852/c
; Sequence 5852, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001

```

```

; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5852
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5852

```

```

Alignment Scores:
Pred. No.: 82.3 Length: 1893
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-489-039A-5852 (1-1893)

```

QY 122 ArgLeuAlaSerSerTyrIleAla 129
DB 1433 CGGTGGCGTCGTCATACATCGCC 1410

```

#### RESULT 8

```

US-09-489-039A-5698
; Sequence 5698, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5698
; LENGTH: 2595
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5698

```

```

Alignment Scores:
Pred. No.: 112 Length: 2595
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 4 Gaps: 0

```

US-09-701-674A-23 (1-179) x US-09-489-039A-5698 (1-2595)

```

QY 122 ArgLeuAlaSerSerTyrIleAla 129
DB 563 CGGTGGCGTCGTCATACATCGCC 586

```

#### RESULT 9

```

US-08-793-824-1
; Sequence 1, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street

```

```

; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4853 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
; US-08-793-824-1

Alignment Scores:
Pred. No.: 204 Length: 4853
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.47% Indels: 0
DB: 2 Gaps: 0

US-09-701-674A-23 (1-179) x US-08-793-824-1 (1-4853)

Qy 165 LysGluValValThrAlaSerArg 172
Db 3241 AAAGAAAGTGTACAGCTTCGTCGT 3264

RESULT 10
US-09-443-199C-469/C
; Sequence 469, Application US/09443199C
; Patent No. 6670464
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin
; TITLE OF INVENTION: Nucleic Acids Containing Single Nucleotide
; FILE REFERENCE: 15966-534A
; CURRENT APPLICATION NUMBER: US/09/443,199C
; CURRENT FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/109,024
; PRIOR FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 1272
; SOFTWARE: CuraGen Patent Formatter Version 0.9
; SEQ ID NO 469
; LENGTH: 51
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (26)...(0)
; OTHER INFORMATION: 1 of 2 allelic variants (470 is other entry)
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)

```

```

; OTHER INFORMATION: Accession number cg32177197
US-09-443-199C-469

Alignment Scores:
Pred. No.: 28.6 Length: 51
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-443-199C-469 (1-51)

Qy 169 ThrAlaSerArgLeuCysGly 175
Db 39 ACGCCTCCAGATTATGTGGA 19

RESULT 11
US-09-237-712-91
; Sequence 91, Application US/09237712
; Patent No. 6180391
; GENERAL INFORMATION:
; APPLICANT: BROWN, WILLIAM C.
; TITLE OF INVENTION: HIGHLY EFFICIENT CONTROLLED EXPRESSION OF EXOGENOUS
; FILE REFERENCE: A-518
; CURRENT APPLICATION NUMBER: US/09/237,712
; CURRENT FILING DATE: 1999-01-26
; EARLIER APPLICATION NUMBER: 60/072,794
; EARLIER FILING DATE: 1998-01-28
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 88
; TYPE: DNA
; ORGANISM: oligonucleotide
US-09-237-712-91

Alignment Scores:
Pred. No.: 48.3 Length: 88
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.91% Indels: 0
DB: 3 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-237-712-91 (1-88)

Qy 84 AlaAsnAlaArgGluArgAla 90
Db 43 GCTAACCGCAGAGAGAGGGCA 63

RESULT 12
US-09-513-999C-14487/C
; Sequence 14487, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14487
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens

```

Search completed: December 17, 2004, 03:14:57  
Job time : 92 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 01:22:13 ; Search time 488 Seconds

(without alignments)  
2025.061 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179

Sequence: 1 MTGSLSDVEDLQEVEMLEC.....PRSDLKVVYTABRLCGTTAS 179

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4093002 seqs, 2760418825 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8176338

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-O=/cgn2\_1/USPTO\_spool\_h/US09701674/runat\_15122004\_100533\_28409/app\_query.fasta\_1.327  
-DS=Published Applications NA -QMT=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -CUTFMF=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USRP=US09701674 @CGN 1.1 480 @runat\_15122004\_100533\_28409  
-NCFU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	52.5	482	10	US-09-918-995-9069
2	61	34.1	697	9	US-09-954-456-829
3	61	34.1	697	9	US-09-954-456-1245
4	61	34.1	697	10	US-09-960-706-1028
5	61	34.1	697	10	US-09-873-319-679
6	49	27.4	632	16	US-10-264-049-1307
7	36	20.1	240	16	US-10-305-720-1028
8	28	15.6	493	10	US-09-918-995-1181
9	28	15.6	1716	17	US-10-775-169-201
10	28	15.6	1939	17	US-10-755-889-365
11	28	15.6	2177	15	US-10-247-671-97
12	28	15.6	2196	15	US-10-084-817-120
13	28	15.6	2382	15	US-10-240-965-83
14	18	10.1	65	10	US-09-908-975-134
15	9	5.0	730	9	US-09-939-825-11
16	9	5.0	1140	17	US-10-450-826-124
17	9	5.0	6288	15	US-10-251-186-23
18	9	5.0	6288	15	US-10-291-172-171
19	9	5.0	6288	16	US-10-221-278-171
20	8	4.5	201	17	US-10-741-601-15181
21	8	4.5	275	11	US-09-955-215-81
22	8	4.5	302	9	US-09-983-965-5366
23	8	4.5	331	16	US-10-424-599-137725
24	8	4.5	369	17	US-10-437-963-76590
25	8	4.5	405	9	US-09-983-965-5418
26	8	4.5	448	16	US-10-425-115-60013
27	8	4.5	453	16	US-10-062-674-910
28	8	4.5	518	18	US-10-363-345A-38647
29	8	4.5	538	16	US-10-363-345A-38648
30	8	4.5	558	10	US-09-918-995-27048
31	8	4.5	565	17	US-10-437-963-37962
32	8	4.5	612	16	US-10-424-599-42408
33	8	4.5	612	16	US-10-425-114-20037
34	8	4.5	633	9	US-09-879-536-493
35	8	4.5	683	13	US-10-027-632-160543
36	8	4.5	683	13	US-10-027-632-160543
37	8	4.5	705	16	US-10-425-114-33618
38	8	4.5	735	13	US-10-027-632-21784
39	8	4.5	735	15	US-10-027-632-21784
40	8	4.5	749	18	US-10-425-115-118396
41	8	4.5	839	18	US-10-425-115-120281
42	8	4.5	840	18	US-10-425-115-85094
43	8	4.5	926	17	US-10-437-963-95779
44	8	4.5	1008	16	US-10-424-599-98612
45	8	4.5			

# ALIGNMENTS

RESULT 1  
US-09-918-995-9069  
; Sequence 9069, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; TITLE OF INVENTION: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9069  
; LENGTH: 482  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

LOCATION: (1).....(482)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-9069

Alignment Scores:  
Pred. No.: 5,76e-89 Length: 482  
Score: 94.00 Matches: 94  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 52.51% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-918-995-9069 (1-482)

QY 86 AlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThr 105  
DB 39 GCGCGAGAGCGGCGCGCGATGCGAGTGTGACGAGGCGCTTCTCCAGACTCAAGACCACC 98  
QY 106 LeuProTrrValProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
DB 99 CTGCCCTGGGTGCCCCCGACACCAAGCTCTCCAAAGCTGGACACGCTCAGCGTGGCGTCC 158  
QY 126 SerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIle 145  
DB 159 AGTACATCGCCACCTTGAGGCGAGATCTTGGCTACGACAAATACGAGACGGGTACATT 218  
QY 146 HisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAspLeuLys 165  
DB 219 CACCCGGTCACTGACGTGGCCCTTATGTGGCGGGAACCCGAGAGTGAACCTGAA 278  
QY 166 GluValValThrAlaSerArgLeuGlyThrThrAlaSer 179  
DB 279 GAAGTGTGACCGGAGCGCGCTTATGTGAACCAACCGCGTCC 320

RESULT 2

US-09-954-456-829  
Sequence 829, Application US/09954456  
Patent No. US20020115057A1  
GENERAL INFORMATION:  
APPLICANT: Young, Paul  
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
FILE REFERENCE: 689290-76  
CURRENT FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US/09/954,456  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/233,617  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/234,052  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,923  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,134  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,637  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,638  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,711  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,720  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,840  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,863  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 2276  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 829  
LENGTH: 697  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
US-09-954-456-829

OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-829

Alignment Scores:  
Pred. No.: 3,35e-54 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-954-456-829 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
DB 3 GACACGCTCAGGCTGGCGTCCAGTACATCGCCCACTTGAGGAGATCTCGGTACCGAC 62  
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
DB 63 AAATACGAGACGGGTACATTACCCGGTCAACCTGACGTCGCGCTTTATGTGGCGCG 122  
QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuGlyThrThrAla 178  
DB 123 AAACCCGAGAGTGAACCTGAAAGAGTGGTGACCGCGAGCGCTTATGTGGAACCAACCGCG 182  
QY 179 Ser 179  
DB 183 TCC 185

RESULT 3

US-09-954-456-1245  
Sequence 1245, Application US/09954456  
Patent No. US20020115057A1  
GENERAL INFORMATION:  
APPLICANT: Young, Paul  
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
FILE REFERENCE: 689290-76  
CURRENT FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: US/09/954,456  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/233,617  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/234,052  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,923  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,134  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,637  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,638  
PRIOR FILING DATE: 2000-09-26  
PRIOR APPLICATION NUMBER: US/60/235,711  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,720  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,840  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: US/60/235,863  
PRIOR FILING DATE: 2000-09-27  
NUMBER OF SEQ ID NOS: 2276  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1245  
LENGTH: 697  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
US-09-954-456-1245

Alignment Scores:  
Pred. No.: 3,35e-54 Length: 697



Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-954-456-1245 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTCTGGCTAACGAC 62  
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AAATACGAGAACGGGTACATTCACCGCGTCAACCTGACGTGGCCCTTTATGGTGGCCGG 122  
QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGAGCCGCTTATGTGGAAACCCCGG 182  
QY 179 Ser 179  
Db 183 TCC 185

RESULT 4

US-09-960-706-1028  
; Sequence 1028, Application US/09960706  
; Publication No. US20030134280A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia  
; FILE REFERENCE: 44921-5029-01US  
; CURRENT APPLICATION NUMBER: US/09/960,706  
; PRIOR FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 60/223,323  
; EARLIER FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 09/873,319  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1028  
; LENGTH: 697  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 W73859  
; NAME/KEY: unsure  
; LOCATION: (1)..(697)  
; OTHER INFORMATION: n = a or c or g or t

US-09-960-706-1028

Alignment Scores:  
Pred. No.: 3.35e-54 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-960-706-1028 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTCTGGCTAACGAC 62  
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AAATACGAGAACGGGTACATTCACCGCGTCAACCTGACGTGGCCCTTTATGGTGGCCGG 122  
QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGAGCCGCTTATGTGGAAACCCCGG 182

QY 179 Ser 179  
Db 183 TCC 185

RESULT 5

US-09-873-319-679  
; Sequence 679, Application US/09873319A  
; Publication No. US20030134324A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; APPLICANT: Kulkarni, Prakash  
; APPLICANT: Getzenberg, Robert H.  
; APPLICANT: Waga, Iwao  
; APPLICANT: Yamamoto, Jun  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia Using Gene Expression Profiles  
; FILE REFERENCE: 44921-5029-US  
; CURRENT APPLICATION NUMBER: US/09/873,319A  
; EARLIER FILING DATE: 2001-06-05  
; EARLIER APPLICATION NUMBER: US 60/223,323  
; EARLIER FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 755  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 679  
; LENGTH: 697  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 W73859  
; NAME/KEY: unsure  
; LOCATION: (1)..(697)  
; OTHER INFORMATION: n = a or c or g or t

US-09-873-319-679

Alignment Scores:  
Pred. No.: 3.35e-54 Length: 697  
Score: 61.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.08% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-873-319-679 (1-697)

QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
Db 3 GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTCTGGCTAACGAC 62  
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AAATACGAGAACGGGTACATTCACCGCGTCAACCTGACGTGGCCCTTTATGGTGGCCGG 122  
QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGAGCCGCTTATGTGGAAACCCCGG 182  
QY 179 Ser 179  
Db 183 TCC 185

RESULT 6

US-10-264-049-1307  
; Sequence 1307, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA133p1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569

1

1

GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES

FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 11181

LENGTH: 493  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(493)  
OTHER INFORMATION: n = A,T,C or G

US-09-918-995-11181

Alignment Scores:  
Pred. No.: 1,01e-19 Length: 493  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-918-995-11181 (1-493)

QY 152 TtpProPheMetValIaGlyLysProGluSerAspLeuLysCluValValThralaser 171  
Db 102 TGCCCTTTATGTGCGCGGAACCCGAGGTGACCTGAAAGAGTGTGACCGGAGC 161  
QY 172 ArgLeuCysGlyThrThralaser 179  
Db 162 CGCTTATGTGGAACCCGCTCC 185

#### RESULT 9

US-10-775-169-201  
Sequence 201, Application US/10775169  
Publication No. US20040175743A1

GENERAL INFORMATION:  
APPLICANT: Wyeth  
APPLICANT: Burczynski, Michael  
APPLICANT: Twine, Natalie  
APPLICANT: Dörner, Andrew  
APPLICANT: Trepicchio, William  
TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
FILE REFERENCE: AM101080 (031896-013000)  
CURRENT APPLICATION NUMBER: US/10/775,169  
CURRENT FILING DATE: 2004-02-11  
NUMBER OF SEQ ID NOS: 5278  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 201  
LENGTH: 1716  
TYPE: DNA  
ORGANISM: Homo sapiens

Alignment Scores:  
Pred. No.: 3.17e-19 Length: 1716  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 17 Gaps: 0

US-10-775-169-201

QY 106 LeuProTropValProPheThrLysLeuSerLysLeuAspThrLeuArgLeuAlaser 125

Db 386 CTGCCCTGGTGGCCCCCGACACTAAGCTCTCCAAGCTGGACACGCTCCGGCTGCC 445  
QY 126 SerTyrlleAlaHisLeuArgGln 133  
Db 446 AGTTACATCGCTCACCTGGCGCAG 469

#### RESULT 10

US-10-755-889-365  
Sequence 365, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755,889  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14  
PRIOR APPLICATION NUMBER: U.S. 60/469,757  
PRIOR FILING DATE: 2003-05-12  
NUMBER OF SEQ ID NOS: 823  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 365  
LENGTH: 1939  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-755-889-365

Alignment Scores:  
Pred. No.: 3.54e-19 Length: 1939  
Score: 28.00 Matches: 28  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.64% Indels: 0  
DB: 17 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-755-889-365 (1-1939)

QY 106 LeuProTropValProPheThrLysLeuSerLysLeuAspThrLeuArgLeuAlaser 125  
Db 612 CTGCCCTGGTGGCCCCCGACACTAAGCTCTCCAAGCTGGACACGCTCCGGCTGCC 671

QY 126 SerTyrlleAlaHisLeuArgGln 133  
Db 672 AGTTACATCGCTCACCTGGCGCAG 695

#### RESULT 11

US-10-247-671-97  
Sequence 97, Application US/10247671  
Publication No. US20030194721A1  
GENERAL INFORMATION:  
APPLICANT: Mikita, Thomas  
APPLICANT: Shiffman, Dov  
APPLICANT: Porter, Gordon, J.  
APPLICANT: Kaser, Matthew R.  
TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
FILE REFERENCE: PA-0050 US  
CURRENT APPLICATION NUMBER: US/10/247,671  
CURRENT FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: 60/323,784  
PRIOR FILING DATE: 2001-09-19  
NUMBER OF SEQ ID NOS: 186  
SOFTWARE: PERL Program  
SEQ ID NO 97  
LENGTH: 2177  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030194721A1 3282941CB1  
US-10-247-671-97

Alignment Scores:  
 Pred. No.: 3,94e-19 Length: 2177  
 Score: 28.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.64% Indels: 0  
 DB: 15 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-247-671-97 (1-2177)

Qy 106 LeuProTnpValProPaspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 857 CTGCCCTGGTGGTCCCGCCGACACTAAGCTCTCCAGCTGGACACGCTCGGCTGGCTTCC 916  
 Qy 126 SerTyrIleAlaHisLeuArgGln 133  
 Db 917 AGTTACATCGCTCACCTGCGGCAG 940

## RESULT 12

US-10-084-817-120  
 ; Sequence 120, Application US/10084817  
 ; Publication No. US20030119009A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Susan Stuart  
 ; APPLICANT: Jed G. Nuchtern  
 ; APPLICANT: Sharon E. Pion  
 ; APPLICANT: Jason M. Shohet  
 ; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION  
 ; FILE REFERENCE: PA-0046 US  
 ; CURRENT APPLICATION NUMBER: US/10/084,817  
 ; CURRENT FILING DATE: 2002-02-25  
 ; PRIOR APPLICATION NUMBER: 60/270,784  
 ; PRIOR FILING DATE: 2001-02-23  
 ; NUMBER OF SEQ ID NOS: 365  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 120  
 ; LENGTH: 2196  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20030119009A1 3282941CB1  
 US-10-084-817-120

Alignment Scores:  
 Pred. No.: 3,97e-19 Length: 2196  
 Score: 28.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.64% Indels: 0  
 DB: 15 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-084-817-120 (1-2196)

Qy 106 LeuProTnpValProPaspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 857 CTGCCCTGGTGGTCCCGCCGACACTAAGCTCTCCAGCTGGACACGCTCGGCTGGCTTCC 916  
 Qy 126 SerTyrIleAlaHisLeuArgGln 133  
 Db 917 AGTTACATCGCTCACCTGCGGCAG 940

## RESULT 13

US-10-240-965-83  
 ; Sequence 83, Application US/10240965  
 ; Publication No. US20030165924A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: INCYTE GENOMICS, INC.  
 ; APPLICANT: SHIFFMAN, Dov  
 ; APPLICANT: SOMOGYI, Roland  
 ; APPLICANT: LAWN, Richard M.  
 ; APPLICANT: SEILHAMER, Jeffrey J.  
 ; APPLICANT: PORTER, Gordon J.

; APPLICANT: MIKITA, Thomas  
 ; APPLICANT: TAI, Julie  
 ; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION  
 ; FILE REFERENCE: PA-0025 PCT  
 ; CURRENT APPLICATION NUMBER: US/10/240,965  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: 60/195,106  
 ; PRIOR FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 276  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 83  
 ; LENGTH: 2382  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20030165924A1 977552.1  
 US-10-240-965-83

Alignment Scores:  
 Pred. No.: 4,27e-19 Length: 2382  
 Score: 28.00 Matches: 28  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.64% Indels: 0  
 DB: 15 Gaps: 0

US-09-701-674A-23 (1-179) x US-10-240-965-83 (1-2382)

Qy 106 LeuProTnpValProPaspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSer 125  
 Db 861 CTGCCCTGGTGGTCCCGCCGACACTAAGCTCTCCAGCTGGACACGCTCGGCTGGCTTCC 920  
 Qy 126 SerTyrIleAlaHisLeuArgGln 133  
 Db 921 AGTTACATCGCTCACCTGCGGCAG 944

## RESULT 14

US-09-908-975-134  
 ; Sequence 134, Application US/09908975  
 ; Publication No. US20030165843A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SHOSHAN, Avi  
 ; APPLICANT: WASSERMAN, Alon  
 ; APPLICANT: MINTZ, Eli  
 ; APPLICANT: MINTZ, Liat  
 ; APPLICANT: FAIGLER, Simchon  
 ; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE  
 ; FILE REFERENCE: 36688-0005  
 ; CURRENT APPLICATION NUMBER: US/09/908,975  
 ; CURRENT FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: US 60/287,724  
 ; PRIOR FILING DATE: 2001-05-02  
 ; PRIOR APPLICATION NUMBER: US 60/221,607  
 ; PRIOR FILING DATE: 2000-07-28  
 ; NUMBER OF SEQ ID NOS: 32337  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 134  
 ; LENGTH: 65  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-09-908-975-134

Alignment Scores:  
 Pred. No.: 4,92e-10 Length: 65  
 Score: 18.00 Matches: 18  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 10.06% Indels: 0  
 DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-908-975-134 (1-65)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeu 18  
 Db 10 ATGTCCACTGGCTCCCTCAGCGATGTAGAAGACCTTCAAGAGGTGGAGATGCTG 63

RESULT 15

US-09-939-825-11/c  
 ; Sequence 11, Application US/09939825  
 ; Patent No. US20020151009A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ni et al.  
 ; TITLE OF INVENTION: PT046P1  
 ; FILE REFERENCE: Human Polynucleotides, Polypeptides, and Antibodies  
 ; CURRENT APPLICATION NUMBER: US/09/939,825  
 ; CURRENT FILING DATE: 2001-08-28  
 ; PRIOR APPLICATION NUMBER: PCT/US01/05498  
 ; PRIOR FILING DATE: 2001-02-22  
 ; PRIOR APPLICATION NUMBER: 60/184,664  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: 60/189,874  
 ; PRIOR FILING DATE: 2000-03-16  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 730  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-939-825-11

Alignment Scores:  
 Pred. No.: 12.4 Length: 730  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 5.03% Indels: 0  
 DB: 9 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-939-825-11 (1-730)

Qy 54 GlyGlyLeuGlyLysArgArgLysala 62  
 Db 128 GGGGGCTTGGGAGAGAGAGAGAGCA 102

Search completed: December 17, 2004, 04:20:36  
 Job time : 549 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 17, 2004, 01:03:08 ; Search time 2869 Seconds  
(without alignments)  
2273.514 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 179  
Sequence: 1 MTGSLSDVEDLQVEMLEC.....PESDLKVVTVASRLCGTTAS 179

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1  
Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool\_h/US09701674/runat\_15122004\_100530\_28319/app\_query.fasta\_1.327  
-DB=EST -Qfmt=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=QUALITY -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09701674 @CNG\_1\_1\_3437@runat\_15122004\_100530\_28319 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	179	100.0	782	6	CF128484
2	179	100.0	796	5	BX107489
3	179	100.0	1044	1	AL550251
4	179	100.0	1221	3	CR594669
5	179	100.0	1250	3	CR606057
6	179	100.0	1252	3	CR616308
7	178	99.4	1038	4	BM544256
8	177	98.9	800	6	CF127519
9	162	90.5	1099	1	AL552119

10	160	89.4	965	1	AL552380
11	153	85.5	1098	4	BM923344
12	151	84.4	757	6	CF126927
13	149	83.2	746	4	BI770998
14	138	77.1	645	6	CF128246
15	138	77.1	744	6	CF128914
16	138	77.1	785	6	CF127896
17	138	77.1	794	6	CF127034
18	136	76.0	738	6	CF130905
19	136	76.0	852	7	CF994025
20	134	74.9	996	5	BX360244
21	133	74.3	827	5	BX926295
22	133	74.3	877	5	CK770542
23	129	72.1	724	6	CF125244
24	127	70.9	1167	4	BM811192
25	125	69.8	563	1	AV605533
26	124	69.3	926	4	BI769090
27	120	67.0	1012	6	BY710826
28	120	67.0	1238	3	AK011575
29	118	65.9	526	1	AJ650722
30	117	65.4	778	7	CO564290
31	117	65.4	818	4	BI767574
32	116	64.8	712	5	BX922447
33	115	64.2	726	5	BX920031
34	112	62.6	629	6	CB447077
35	111	62.0	651	4	BI765033
36	110	61.5	589	6	CB417689
37	108	60.3	605	4	BI540239
38	108	60.3	784	7	CK770983
39	108	60.3	1085	1	AL574170
40	107	59.8	717	6	CF126467
41	107	59.8	1065	5	BQ072411
42	102	57.0	1169	1	AL552084
43	100	55.9	615	2	AW920010
44	100	55.9	1024	1	AL575799
45	98	54.7	597	5	BU358809

## ALIGNMENTS

RESULT 1  
CF128484

UI-HF-ETO-awj-c-15-0-UI.r1 NIH\_MGC\_214 Homo sapiens CDNA clone  
IMAGE:30557534 5', mRNA sequence.

CF128484 782 bp mRNA linear EST 05-AUG-2003  
CF128484.1 GI:33207776

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Normalizaton and subtraction: two approaches to facilitate gene  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Mary Hendrix  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/humanfl.html>  
The following repetitive elements were found in this CDNA

sequence: 1-81, >(GA)n#simple\_repeat (matched complement)

Seq primer: PYX-5

# FEATURES

source Location/Qualifiers

1..782  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30557534"  
 /tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_MGC\_214"  
 /notes="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATPAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,01e-169 Length: 782  
 Score: 179.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128484 (1-782)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 86 ATGTCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTGGAGATGTGGAATGT 145  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 DB 146 GAGCGGTTGAAATGATTCGAAACAGGAATTTGTGACTTCCACGAGCAGCACCAGGAG 205  
 QY 41 SerSerAsnGlySerProGlnLysGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60  
 DB 206 AGTCTCAACTGCGAAGATGGGTCTCCCAAGAGGGCCGCGCGCTGGCCAGAGGAGG 265  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 DB 266 AAGCGCCCCACCAAGAAGAGCCCTCAGCGGGTCCAGCCAGGAGGGAAGCAGGTCCAG 325  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 DB 326 CGCAACGCGCCCAACCGCGAGAGCGGCCCGCATGCGAGTGTGAGCAAGGCCCTTCCTCC 385  
 QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 DB 386 AGACTCAAGACCACTCTGCGCTGGTGGTCCCGCCGACACCACTCTCCAGCTGACACG 445  
 QY 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysThr 140  
 DB 446 CTCAGCTGGCGTCCAGCTACATCGCCCACTTGGAGCAGATCTCGCTGCAACGCAATAC 505  
 QY 141 GluAsnGlyThrIleHisProValAsnLeuThrTrpPheMetValAlaGlyLysPro 160  
 DB 506 GAGAACGGGTACATTCACCGGTCACTGAGCTGGCCCTTTATGGTGGCCGGGAACCC 565  
 QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 DB 566 GAGAGTGACCTGAAAGAGTGTGTGACCGGAGCGGCTTATGTGGAACCAACCGCGTCC 622

## RESULT 2

BX107489

LOCUS

796 bp mRNA

linear

EST 06-FEB-2003

## DEFINITION

BX107489 Soares placenta Nb2HP Homo sapiens cDNA clone  
 IMAGE998N24234 ; IMAGE:151511, mRNA sequence.

## ACCESSION

BX107489

## VERSION

BX107489.1 GI:27834707

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## REFERENCE

1 (bases 1 to 796)

## AUTHORS

Eberly, L., Heil, O., Hennig, S., Neubert, P., Peters, M., Radelof, J., Schneider, D. and Korn, B.

## TITLE

Human Unigeneset - RZPD3

## JOURNAL

Unpublished (2003)

## COMMENT

Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD; IMAGE998N24234.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Human Unigeneset - RZPD3 (RZPDLIB No.972).

http://www.rzpd.de/CloneCards/cgi-

bin/showLib.pl.cgi/?response?libNo=972 Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32839 101

Fax: +49 30 32839 111

www.rzpd.de

This clone is available royalty-free from RZPD;

contact RZPD (clone@rzpd.de) for further information. Seq primer:

M13r, Primer sequence: TTTCACAGGAACAGCTATGAC.

## FEATURES

Location/Qualifiers

1..796

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE998N24234 ; IMAGE:151511"

/sex="Female"

/dev\_stage="Placenta obtained at birth (full term)"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_lib="Soares placenta Nb2HP"

/notes="Organ: placenta; Vector: pYX-Asc (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer (5'

AAGTGAAGAATTCGGCGGCGGAGGATTTTTTTTTTTT 3').

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pYX-Asc vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo. "

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,06e-169 Length: 796  
 Score: 179.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 5 Gaps: 0

US-09-701-674A-23 (1-179) x BX107489 (1-796)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 DB 195 ATGTCCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAGAGGTGGAGATGTGGAATGT 254  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 DB 255 GACGGGTGAAATGGAATTCGAACAGGAATTTGTGACTTCCACGAGACCCAGGAG 314  
 QY 41 SerSerAsnGlySerProGlnLysGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 60  
 DB 315 AGTCCCAACTGCCAGATGGGTCTCCCAAGAGGGCCGCGGCGCTCGGCAAGAGGAG 374







```

KEYWORDS      HTC; CNSLT_CDNA.
SOURCE         Homo sapiens (human)
ORGANISM       Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1252)
AUTHORS       Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE         Full-length cDNA libraries and normalization
JOURNAL       Unpublished
REMARK        Contact : Feng Liang Email : fliang@lifetech.com URL :
               http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
               Faraday Avenue
REFERENCE     2 (bases 1 to 1252)
AUTHORS       Genoscope.
TITLE         Direct Submission
JOURNAL       Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
               BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
               - Web : www.genoscope.cns.fr)
COMMENT       1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned
               into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by Life Technologies, a
               division of Invitrogen.
FEATURES      Location/Qualifiers
               1..1252
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS01070YK16"
                /issue_type="Placenta Cot 25-normalized"
                /plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.:      4,64e-169      Length:      1252
Score:          179.00         Matches:    179
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    100.00%       Indels:      0
DB:             3              Gaps:       0

US-09-701-674A-23 (1-179) x CR616308 (1-1252)
QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db      272 ATGTCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTGGAATGT 331
QY      21 AspGlyLeuLeuMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      332 GACGGGTTCGAAATGGATTTCGAACAAGGAATTTGTGACTTCCACGAGCAGCAGCAGGAG 391
QY      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60
Db      392 AGCTCCAACTGGAGATGGGTCTCCCAAGAGGGCGCGGCGCTGGGCAAGAGGAGG 451
QY      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      452 AAGCGCGCCCAACAAGAGAGCCCTCAGCGGGGTGAGCCAGAGGGGGAAGCAGGTCCAG 511
QY      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      512 CGCAACGCCGCCACACGCCGAGAGGGGCCCGCCATCGGAGTCTCAGCAAGGCCCTTCTCC 571
QY      101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120
Db      572 AGACTCAAGACCAACCCCTGGCTGGGTGCGCCCGCACCAACCACTCTCCAAGCTGGACACG 631
QY      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      632 CTCAGCTGGCTTCAGCTACATCGCCCATTTGAGCGAGATCCTGGCTTAACGACAAATAC 691
QY      141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160

```

---

```

Db      692 GAGAACGGGTACATTCACCCGGTCAACCTGACGTGGCCCTTTATGGTGGCGGGAACCC 751
QY      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db      752 GAGAGTGCCTGAAGAGAGTGGTGACCGCGCGCTTATGTGGAACACCGCGCTCC 808

RESULT 7
BM544256      1038 bp      mRNA      linear      EST 20-FEB-2002
LOCUS        AGENCOURT 6490671 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5587760
5', mRNA sequence.
ACCESSION    BM544256
VERSION      BM544256.1 GI:18775358
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 1038)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Invitrogen
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM12357 row: m column: 09
              High quality sequence stop: 774.
FEATURES      Location/Qualifiers
               1..1038
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5587760"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_125"
                /notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
                of three ovaries, from females ranging in age from 38 to
                49 yo. Library is oligo-dT primed and directionally cloned
                (EcoRV site is destroyed upon cloning). Average insert
                size 2.1 kb, insert size range 1-3.5 kb. Library is
                normalized and enriched for full-length clones and was
                constructed by C. Gruber (Invitrogen). Research Genetics
                tracking code 036."
ORIGIN
Alignment Scores:
Pred. No.:      3,95e-168      Length:      1038
Score:          178.00         Matches:    178
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    99.44%       Indels:      0
DB:             4              Gaps:       0

US-09-701-674A-23 (1-179) x BM544256 (1-1038)
QY      2 SerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAsp 21
Db      82 TCCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTGGAATGTGAC 141
QY      22 GlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSer 41
Db      142 GGGTGTGAAATGATTCGAACAAGGAATTTGTGACTTCCACGAGACACCGAGGAGAC 201
QY      42 SerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLysArgArgLys 61
Db      202 TCCAACCTGCAGATGGTCTCCCAAGAGGGCGCGGCGCTGGCAAGAGGAGGAG 261

```

gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,15e-167 Length: 800  
 Score: 177.00 Matches: 177  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.88% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674a-23 (1-179) x CF127519 (1-800)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 258 ATGTCCACCGGCTCCCTCAGCATGTGGAGACCTTCAAGAGGTGGAGATGTTGGAATGT 317  
 Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 318 GACGGGTGTAATGGAATTCGAACAAGGAATTTGTGACTTCCAAAGAGACCGGAGGAG 377  
 Qy 41 SerSerAsnGlyGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
 Db 378 AGTCCCACTGCCAGATGGGTCTCCCAAGAGGCGCGGGCGCTGGCAAGAGGAGG 437  
 Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGluLysGlnValGln 80  
 Db 438 AAGCGGCCCAACCAAGAGAGCCCTCAGCGGGGTGAGCCAGGAGGAGGAGGAGTCCAG 497  
 Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 498 CGCAACGCCGCCCAACCGCGAGAGCGGGCCGCGATGCGAGTGTGAGCAAGGCTTCTCC 557  
 Qy 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 558 AGACTCAAGACCAACCTCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 617  
 Qy 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140  
 Db 618 CTGAGGTGGCGTCCAGCTACATCGCCCATCTGAGGAGATCTGCTACGACCAATAC 677  
 Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 678 GAGAACGGGTACATTCACTCCCGGTCACTGACGTGGCCCTTTATGTGGCGGAAACCC 737  
 Qy 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThr 177  
 Db 738 GAGAGTGACCTGAAGAAGTGGTGGCGGAGCGCTTATGTGGAACCCAG 788

## RESULT 9

## AL552119

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AL552119 Homo sapiens 1099 bp mRNA linear EST 30-MAR-2004  
 Clone CSODI059YH17 5-PRIME, mRNA sequence.  
 AL552119  
 AL552119.3 GI:45856908  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1099)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On Feb 15, 2001 this sequence version replaced gi:31273935.  
 Contact: Genoscope

Qy 62 AlaProThrLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArg 81  
 Db 262 GCGCCCAACAAGAGGCGGCTGAGCGGGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 321  
 Qy 82 AsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArg 101  
 Db 322 AAGCGCGCCCAACCGCGAGAGCGGCGCGCATGCGAGTGTGAGCAAGGCTTCTCCAGA 381  
 Qy 102 LeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThrLeu 121  
 Db 382 CTCAGAGCACCGCTGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 441  
 Qy 122 ArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysTyrGlu 141  
 Db 442 AGCTGGCGTCCAGTACATCGCCCACTGAGGCGAGATCTGCTAACGACAAATACAG 501  
 Qy 142 AsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGlu 161  
 Db 502 AACCGGTACATTCAACCGGTCAACCTGAGTGGCGCTTTATGTGGCGGCAAAACCCGAG 561  
 Qy 162 SerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 562 AGTGACCTGAAGAAGTGGTGGCGGAGCGGCTTATGTGGAACCAACCGCGTCC 615

## RESULT 8

## CF127519

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

CF127519 800 bp mRNA linear EST 05-AUG-2003  
 UI-HP-E70-awh-f-14-0-UI.r1 NIH\_MGC\_214 Homo sapiens cDNA clone  
 IMAGE:30554533 5', mRNA sequence.  
 CF127519  
 CF127519.1 GI:33205841  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 800)  
 Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 9704477  
 889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/humanfl.html>  
 The following repetitive elements were found in this cDNA  
 sequence: 172-253, >(G)A#Simple\_repeat (matched complement)  
 Seq primer: pYX-5,  
 Location/Qualifiers  
 1..800  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30554533"  
 /tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
 /lab\_host="DH10B (r1 phage resistant)"  
 /clone\_lib="NIH MGC 214"  
 /note="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose

## FEATURES

## source

Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f

For more information about this cluster, see  
<http://www.genoscope.cns.fr/cdna?s=CSODI059CD09QPlsc=3812.f>.

# FEATURES

source

1..1099  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI059YH17"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.  
 Library was normalized."

# ORIGIN

Alignment Scores:  
 Pred. No.: 5,03e-152 Length: 1099  
 Score: 162.00 Matches: 162  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 90.50% Indels: 0  
 DB: 1 Gaps: 0

US-09-701-674A-23 (1-179) x AL552119 (1-1099)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValMetLeuGluCys 20  
 Db 260 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTTGAATGT 319  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 320 GACGGGTGAAATGGATTTCGAACAGGAATTTGTGACTTCCACGAGAGACCGAGGAG 379  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 Db 380 AGCTCCAACTGCAGAAATGGTCTCCCAAGAGGCGCGCGCTGGCAAGAGGAGG 439  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 440 AAGGCGCCCAACCAAGAGAGCCCTGTAGCGGGGTCCAGCAGAGGGGAAGCAGGTCCAG 499  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 500 CGCAACCGCCCAACCGCGAGAGCGGGCCCGATCGGAGTTCTGAGCAAGGCCCTTCTCC 559  
 QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 560 AGACTCAAGACCAACCTCGCTGGGTGCGCCCGCACCAAGCTCTCCAAAGCTGACACG 619  
 QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 620 CTCAGCTGGCGTCCAGCTACATCGCCCACTTGAAGCAGATCCTGGCTAACACAAATAC 679  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 680 GAGAACGGGTACATTACCGGTCACTGACGTGGCCCTTTATGTTGGCGGGAAACCC 739  
 QY 161 GluSer 162  
 Db 740 GAGAGT 745

RESULT 10  
 AL552380

LOCUS AL552380 965 bp mRNA linear EST 30-MAR-2004  
 DEFINITION AL552380 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
 clone CSODI070YX16 5-PRIME, mRNA sequence.

ACCESSION AL552380

VERSION AL552380.3 GI:45857168

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 965)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT On Feb 15, 2001 this sequence version replaced gi:31274195.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f

For more information about this cluster, see

<http://www.genoscope.cns.fr/cdna?s=CSODI070BF08QPlk=3812.f>.

# FEATURES

source

Location/Qualifiers  
 1..965  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI070YX16"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.  
 Library was normalized."

# ORIGIN

Alignment Scores:  
 Pred. No.: 4,56e-150 Length: 965  
 Score: 160.00 Matches: 175  
 Percent Similarity: 99.43% Conservative: 0  
 Best Local Similarity: 99.43% Mismatches: 1  
 Query Match: 89.39% Indels: 1  
 DB: 1 Gaps: 0

US-09-701-674A-23 (1-179) x AL552380 (1-965)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 272 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTTGAATGT 331  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 332 GACGGGTGAAATGGATTTCGAACAGGAATTTGTGACTTCCACGAGAGACCGAGGAG 391  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 Db 392 AGCTCCAACTGCAGAAATGGTCTCCCAAGAGGCGCGCGCTGGCAAGAGGAGG 451  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 452 AAGCGCGCCCAACCAAGAGAGCCCTCAGCGGGGTCCAGCAGAGGGGAAGCAGGTCCAG 511  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 512 CGCAACCGCCCAACCGCGAGAGCGGGCCCGATCGGAGTTCTGAGCAAGGCCCTTCTCC 571  
 QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120

```

Db      572 AGACTCAAGACACCCCTGCGGTGGTGGCCCGCCAGACACCAAGCTCTCCAAAGCTGGACACG 631
Qy      121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      632 CTAGGCTGGCGTCCAGCTACATCGCCCACTTGAGCGAGATCCTGGCTAAGCAAAATAC 691
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTTPProPheMetValAlaGlyLysPro 160
Db      692 GAGACGGGTACATTCACCGGTCACCTGACGTCGGCCCTTATGTTGGCGGGAAACC- 750
Qy      161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThr 176
Db      751 GAGAGTGACCTGAAGAAGTGTGACCGCGAGCCGCTTATGTGGAACA 798

RESULT 11
LOCUS   BM923344
DEFINITION AGENCOURT_6626079 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5758942
5', mRNA sequence.
ACCESSION BM923344
VERSION   BM923344.1 GI:19373723
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1098)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILLNL at:
http://image.llnl.gov
Plate: LLAM12803 row: i column: 23
High quality sequence stop: 693.
Location/Qualifiers
1..1098
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5758942"
/lab_host="DH10B"
/clone_lib="NIH_MGC_116"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 5 58e-143 Length: 1098
Score: 153.00 Matches: 153
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.47% Indels: 0
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x BM923344 (1-1098)
Qy      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20

```

```

Db      268 ATGTCCACCGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTTGAATGT 327
Qy      21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db      328 GACGGGTGAAATGATTCGAACAAGGAATTTGTGACTTCCAAACGAGACCCGAGGAG 387
Qy      41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60
Db      388 AGCTCCAACTGCCAGAAATGGTCTCCCAAGAGGGCGCGGGCTGGCAAGAGAGG 447
Qy      61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db      448 AAGCGCCCAACCAAGAGAGCCCTTGGAGCGGGGTGAGCCAGAGGGAGAGGTCCTCCAG 507
Qy      81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db      508 CGCAACGCCGCCAAGCGCGAGAGCGGGCCGCGATCGAGTGTCTGAGCAAGGCTTCTCC 567
Qy      101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
Db      568 AGACTCAAGACCACTCCCTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 627
Qy      121 LeuArgGluAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db      628 CTAGGCTGGCGTCCAGCTACATCGCCCACTTGAGCGAGATCCTGGCTAAGCAAAATAC 687
Qy      141 GluAsnGlyTyrIleHisProValAsnLeuThrTTPPro 153
Db      688 GAGACGGGTACATTCAACCGGTCAACCTGACGTGGCCC 726

RESULT 12
LOCUS   CF126927
DEFINITION CF126927 757 bp mRNA linear EST 05-AUG-2003
UI-HF-ET0-avx-k-09-0-UI-r1 NIH_MGC_214 Homo sapiens cDNA clone
IMAGE:30563480 5', mRNA sequence.
ACCESSION CF126927
VERSION   CF126927.1 GI:33204654
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 757)
AUTHORS Ronaldo M.F., Lennon G. and Soares M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Mary Hendrix
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanfl.html
The following repetitive elements were found in this cDNA
sequence: 177-258, >(GA)n#Simple_repeat (matched complement)
Seq primer: pXX-5,
Location/Qualifiers
1..757
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30563480"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
FEATURES
source

```

/lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_MGC\_214"  
 /note="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to Not I site. Double strand cDNA was  
 with EcoR I adaptor, digested with Not I and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is GATAGGCCCA. Tissue was provided by Mary Hendrix."

ORIGIN

Alignment Scores:  
 Pred. No.: 4,06e-141 Length: 757  
 Score: 151.00 Matches: 151  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 84.36% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF126927 (1-757)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 263 ATGTCCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 322  
 QY 21 AspGlyLeuMetAspSerLeuSerGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 323 GACGGGTGAAATGGATTCGACAAAGGAATTTGTGACTTCCAAAGAGACCCGAGGAG 382  
 QY 41 SerSerAsnGlySerProGlnGlyGlyArgGlyGlyGlyLeuGlyLysArgArg 60  
 Db 383 AGCTCAACTGCGAGAAATGGTCTCTCCAGAGAGGCGCGCGCTGGGCAAGAGGAGG 442  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 443 AAGGCGCCCAACCAAGAGAGCGCCCTGAGCGGGTCCAGCCAGGAGGAGGAGTCCAG 502  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 503 CGCAACGCGCCACACGCGGAGGCGGCGCGCATGCGAGTCTGAGCAAGCCCTTCTCC 562  
 QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 563 AGACTCAAGACCAACCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 622  
 QY 121 LeuArgLeuAlaSerSerTyrrileAlaHisLeuArgGlnileAlaAsnAspLysTyrr 140  
 Db 623 CTCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 682  
 QY 141 GluAsnGlyTyrrileHisProValAsnLeuThr 151  
 Db 683 GAGAACGGGTACATTCAACCGGTCAACCTGACG 715

RESULT 13

BI770998  
 LOCUS 60305534:1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5204651 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI770998  
 VERSION BI770998.1 GI:15762576  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 746)  
 NIH-MGC http://mgs.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: rcaphe@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 cDNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM1513 row: b column: 12  
 High quality sequence stop: 744.  
 Location/Qualifiers  
 1..746  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5204651"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_122"  
 /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
 Site 1: NotI; Site 2: EcoRV (destroyed); RNA source  
 anonymous pool of 24 week female lung, 16 week female  
 spleen, and 20-22 week male spleens. Library is oligo-dT  
 primed and directionally cloned (EcoRV site is destroyed  
 upon cloning). Average insert size 1.4 kb, insert size  
 range 1-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 026. Note:  
 this is a NIH\_MGC Library."

FEATURES

source

ORIGIN

Alignment Scores:  
 Pred. No.: 4.1e-139 Length: 746  
 Score: 149.00 Matches: 162  
 Percent Similarity: 99.39% Conservative: 0  
 Best Local Similarity: 99.39% Mismatches: 0  
 Query Match: 83.24% Indels: 1  
 DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x BI770998 (1-746)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 258 ATGTCCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 317  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 318 GACGGGTGAAATGGATTCGACAAAGGAATTTGTGACTTCCAAAGAGACCCGAGGAG 377  
 QY 41 SerSerAsnGlySerProGlnGlyGlyArgGlyGlyGlyLeuGlyLysArgArg 60  
 Db 378 AGCTCAACTGCGAGAAATGGTCTCTCCAGAGGCGCGCGGCTGGGCAAGAGGAGG 437  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 438 AAGCGCCCAACCAAGAGAGCCCTCAGCGGGTCCAGCCAGGAGGAGGAGGAGTCCAG 497  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 498 CGCAACCGCCCAACGCGGAGAGCGGCGCGCATGCGAGTGTGAGCAAGGCGCTTCTCC 557  
 QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 558 AGACTCAAGACCAACCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 617  
 QY 121 LeuArgLeuAlaSerSerTyrrileAlaHisLeuArgGlnileAlaAsnAspLysTyrr 140  
 Db 618 CTCAGGTGGGCTCCAGCTACATCGCCCACTTGGGCGAGATCTCTGGTGAACGACAAATAC 677  
 QY 141 GluAsnGlyTyrrileHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPr 160  
 Db 678 GAGAACGGGTACATTCAACCGGTCAACCTGAGCGGCTTTATGGTGGCGCGGAGAAC 737

```

QY      160 oGluSer 162
Db      738 CGAGAGT 744

RESULT 14
CF128246
LOCUS   645 bp mRNA linear EST 05-AUG-2003
DEFINITION
UI-HF-ET0-awf-c-24-0-UI.r1 NIH_MGC_214 Homo sapiens cDNA clone
IMAGE:30555623 5', mRNA sequence.
ACCESSION
CF128246
VERSION
CF128246.1 GI:33207297
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 645)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Mary Hendrix
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanfl.html
The following repetitive elements were found in this cDNA
sequence: 39-120, >(GA)n#Simple_repeat (matched complement)
Seq primer: pyX-5.

FEATURES
source
Location/Qualifiers
1..645
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30555623"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="NIH_MGC_214"
/clone_lib="NIH_MGC_214"
/notes="Organ: Lung; Vector: pyX-Asc; Site: 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pyX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GATAGGCCA. Tissue was provided by Mary Hendrix."

ORIGIN
Alignment Scores:
Pred. No.: 4,08e-128 Length: 645
Score: 138.00 Matches: 138
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.09% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128246 (1-645)

QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
|||||

```

```

125 ATGTCCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGAGAGTGTGGATGT 184
21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
185 GACGGGTGAAATGGATTCCAAAGGAATTTGTGACTTCCAAACGAGACCCGAGGAG 244
41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLysGlyLysGly 60
245 AGTCTCAACTGCCAGAAATGGGTCTCCCAAGAGGGCGCGCGCTGGCGAAGAGAG 304
61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGlnGlyLysGlnValGln 80
305 AAGGGCGCCCAACCAAGAGAGCCCTGAGCGGGGTGAGCCAGGAGGAGGAGGTCCAG 364
81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
365 CGCAACGCCGCCAACCGCGAGAGCGGGCGCGCATCGAGTGTCTGAGCAAGGCTTCTCC 424
101 ArgLeuLysThrThrLeuProTtpValProAspThrLysLeuSerLysLeuAspThr 120
425 AGACTCAAGACCACTCCCTGGTGGTGGCCCGACCAAGCTCTCCAGCTGGACAG 484
121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138
485 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCTCGCTAACGCAC 538

RESULT 15
CF128914
LOCUS   744 bp mRNA linear EST 05-AUG-2003
DEFINITION
UI-HF-ETO-avu-p-09-0-UI.r1 NIH_MGC_214 Homo sapiens cDNA clone
IMAGE:30563216 5', mRNA sequence.
ACCESSION
CF128914
VERSION
CF128914.1 GI:33208641
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 744)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Mary Hendrix
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/humanfl.html
The following repetitive elements were found in this cDNA
sequence: 191-272, >(GA)n#Simple_repeat (matched complement)
Seq primer: pyX-5.

FEATURES
source
Location/Qualifiers
1..744
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30563216"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="NIH_MGC_214"
/clone_lib="NIH_MGC_214"
/notes="Organ: Lung; Vector: pyX-Asc; Site: 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with Not I and then cloned
directionally into pyX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GATAGGCCA. Tissue was provided by Mary Hendrix."

ORIGIN
Alignment Scores:
Pred. No.: 4,08e-128 Length: 645
Score: 138.00 Matches: 138
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.09% Indels: 0
DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x CF128246 (1-645)

QY      1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
|||||

```



19%. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

## Alignment Scores:

Pred. No.:	4.65e-128	Length:	744
Score:	138.00	Matches:	138
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	77.09%	Indels:	0
DB:	6	Gaps:	0

US-09-701-674A-23 (1-179) x CF128914 (1-744)

QY	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys	20
DB	277	ATGTCACCGCCCTCCCTGAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT	336
QY	21	AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
DB	337	GACGGGTTGAAATGATTCGAACAGGAATTTGTGACTTCCACGAGAGCACCAGGAG	396
QY	41	SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg	60
DB	397	AGCTCCAACTGGAGAATCGGTCTCCCCAGAAAGGGCCGCGCGCTGGGCAAGAGAGG	456
QY	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln	80
DB	457	RAGGCGCCCAACAGAGAGGCCCCCTGAGCGGGTCAGCCAGAGGGGAGGAGGTCCAG	516
QY	81	ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer	100
DB	517	CGCAGCGCGCCCAACCGCGAGAGCGGGCCGCGATGCGAGTCTGAGCAAGGCCCTCTCC	576
QY	101	ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr	120
DB	577	AGACTCAGACCAACCCCTGCTGGTGCCCGCCCGACACCAAGCTCTCCAAAGCTGGACAG	636
QY	121	LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp	138
DB	637	CTCAGGCTGGGTCCAGCTACATCGCCCACTTGGAGCAGATCCTGGCTAACGAC	690

Search completed: December 17, 2004, 03:13:22  
Job time : 2874 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 23:16:02 ; Search time 487 Seconds  
(without alignments)  
2029.220 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MSTGSLSDVEDLQEVEMLEC.....PESDLKEVTVASRLGGTTAS 179

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4093002 seqs, 2760418825 residues

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool.h/US09701674/runat\_15122004\_100443\_27616/app.query.fasta\_1.327  
-DB=Published Applications NA -QWMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -THR SCORE=pct -THR MAX=100  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -END=1 -MAIRIX=blosum62  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09701674@cgn 1 1 480 @runat\_15122004\_100443\_27616  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:  
20: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
21: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ORGANISM: Homo sapiens

FEATURE:

Result No.	Score	Match	Length	DB ID	Description
1	764.5	83.4	632	16	US-10-264-049-1307
2	514	56.1	2177	15	US-10-247-671-97
3	514	56.1	2196	15	US-10-084-817-120
4	507	55.3	2382	15	US-10-240-965-83
5	493.5	53.8	1716	17	US-10-775-169-201
6	487	53.1	482	10	US-09-918-995-9069
7	482	52.6	1939	17	US-10-755-889-365
8	353	38.5	240	16	US-10-305-720-1028
9	318	34.7	697	9	US-09-954-456-829
10	318	34.7	697	9	US-09-954-456-1245
11	318	34.7	697	10	US-09-960-706-1028
12	318	34.7	697	10	US-09-873-313-679
13	228	24.9	493	10	US-09-918-995-11181
14	180.5	19.7	1140	17	US-10-450-826-124
15	168.5	18.4	446	9	US-09-833-381-505
16	165.5	18.0	651	9	US-09-749-728B-22
17	165.5	18.0	2465	13	US-10-044-090-251
18	164	17.9	609	15	US-10-029-386-22740
19	164	17.9	1800	10	US-09-771-357-106
20	164	17.9	1800	15	US-10-059-579-106
21	162	17.7	1457	9	US-09-954-531-982
22	162	17.7	1467	16	US-10-062-674-1950
23	162	17.7	1678	15	US-10-240-965-119
24	157.5	17.2	907	13	US-10-004-717-65
25	157.5	17.2	907	18	US-10-860-373-65
26	157.5	17.2	907	18	US-10-860-724-65
27	157.5	17.2	2537	16	US-10-466-164-25
28	157.5	17.2	2550	18	US-10-335-053-106
29	157.5	17.2	2601	9	US-09-880-107-2433
30	157	17.1	1477	13	US-10-004-717-63
31	157	17.1	1477	18	US-10-860-373-63
32	157	17.1	1477	18	US-10-860-724-63
33	151.5	16.5	501	13	US-10-004-717-13
34	151.5	16.5	501	18	US-10-860-373-13
35	151.5	16.5	501	18	US-10-860-724-13
36	151	16.5	515	13	US-10-004-717-35
37	151	16.5	515	18	US-10-860-373-35
38	151	16.5	515	18	US-10-860-724-35
39	150.5	16.4	4153	13	US-10-116-275-194
40	148.5	16.2	993	13	US-10-004-717-47
41	148.5	16.2	993	18	US-10-860-373-47
42	148.5	16.2	993	18	US-10-860-724-47
43	148.5	16.2	3261	13	US-10-004-717-12
44	148.5	16.2	3261	18	US-10-860-373-12
45	148.5	16.2	3261	18	US-10-860-724-12

# ALIGNMENTS

RESULT 1  
US-10-264-049-1307  
; Sequence 1307, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA133p1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: US 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 4360  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 1307  
; LENGTH: 632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

```

/ NAME/KEY: misc_feature
/ LOCATION: (78)..(79)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (174)..(174)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (231)..(231)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (249)..(249)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (263)..(263)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (265)..(265)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (280)..(280)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (331)..(331)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (433)..(433)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (467)..(467)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (498)..(498)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (537)..(537)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (546)..(546)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (552)..(552)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (623)..(623)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (629)..(629)
/ OTHER INFORMATION: n equals a,t,g, or c
/ US-10-264-049-1307

```

US-09-701-674A-23 (1-179) x US-10-264-049-1307 (1-632)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 13 ATGTCACCGGCTCCCTCAGCGATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 72  
Qy 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 73 GACGGNWTGAATGGATTGGACACAGGAATTGTGACTTCCACAGAGACACCAGGAG 132  
Qy 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyClyLeuGlyLysArgArg 60  
Db 133 AGCTCCAACTCGAGAATGGGTCTCCACAGAGGGCGCGGNGGCTGTGGCCAAAGAGGAG 192  
Qy 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
Db 193 AAGGGCCGCCCAAGAGAGAGCCCCCTGAGCGGTGTGACNAGAGGGGAGCAGGTCNAG 252  
Qy 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 253 CGCAACGCCGNCNACGGCGAGAGCGGNCGCCGATCGAGTGTGACGACGAAGGCTTCTCC 312  
Qy 101 ArgLeuLysThrThrLeuProTtpValProAspThrLysLeuSerLysLeuAspThr 120  
Db 313 AGACTCAAGACCAACCTGTCCTGGGTGCCCCCGACCAACGACTCTCCAAAGCTGGACAG 372  
Qy 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 373 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCGAGATCCTGGCTAACGACCAATAC 432  
Qy 141 GluAsnGlyTyrIleHisProValAsnLeuThrTtpProPheMetValAlaGlyLysPro 160  
Db 433 NAGAACGGGTACATTA--CCGGGCAACTGACGTGGNCCTTTATGGTGGC--GGGAACCC 489  
Qy 161 GluSerAspLeuLysGluValVal--ThrAlaSerArgLeuCysGlyThrAla 178  
Db 490 GAGATGANTGAAAGAGTGTGACCGCGAGCCGCTTATGTGGAANCACCGCG 544

RESULT 2  
US-10-247-671-97  
; Sequence 97, Application US/10247671  
; Publication No. US20030194721A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikita, Thomas  
; APPLICANT: Shifman, Dov  
; APPLICANT: Porter, Gordon, J.  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
; FILE REFERENCE: PA-0050 US  
; CURRENT APPLICATION NUMBER: US/10/247,671  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: 60/323,784  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 186  
; SOFTWARE: PERL Program  
; SEQ ID NO 97  
; LENGTH: 2177  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID NO. US20030194721A1 3282941.CB1  
US-10-247-671-97

Qy	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys	20
Db	458	ATGTCCACGGCTCGGTGAGTGATCCGGAGGAG	499
Qy	21	AspGlyLeuLys	Met 25
Db	500	CGGGGCTGCAGCGGGAGTACCCGGTCCCGCCTCCAAGAGCGCGCCCTCCGCGGGGTA	559
Qy	26	AspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
Db	560	GAGCGCAGCTACGCCTCGCCACAGTGACAACCTCGTCGCAGAGGAGGAGGACCCGACGGC	619
Qy	41	-----SerSerAsnCysGluAsnGlySer	48
Db	620	GAGGAGGAGCGCTCGCTCTGGGCACAGCGCGCAGCGCGGAGGCTCCAGAGGAGCGG	679
Qy	49	ProGln---LysGlyArgGlyGlyLeuGlyLysArgGlyLysAlaProThrLysLysSer	67
Db	680	CCCCGTGTGCTGGGGCGCGCGCAGGTGGTAGCGCGCGCGGTGTGCACAGAAGCC	739
Qy	68	---ProLeuSerGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAla	86
Db	740	CTCCCGGCCAAGGGCTCAGCGCGCAGATGCAGAGCTCGCAGCGGACGGCGCCAGCC	799
Qy	87	ArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeu	106
Db	800	CGTGAGCGTGCCCGGATCGCGTGCTCAGCAAGCCTTCTCCAGGCTCAAGACGACGCTG	859
Qy	107	ProTyrValProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSer	126
Db	860	CCCTGGGTGCCCCCGACATTAGCTCTCCAGCTGGACACGCTCCGGCTGGCTTCAGT	919
Qy	127	TyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIleHis	146
Db	920	TACATCGCTCACCTCGCGCAGCTGTTCCAGGAGGACCGCTATGAGAAGCGCTACGTGCAC	979
Qy	147	ProValAsnLeuThrTyrProPheMetValAlaGlyLysProGluSerAspLeuLysGlu	166
Db	980	CCAGTGAACCTGACATGCCCATTCGTGGTCTCGGGAAGCCGGACTCTGACACCAAGAA	1039
Qy	167	ValValThrAlaSerArgLeuCysGlyThrThrAla	178
Db	1040	GTITCCCGACCAACAGACTATGTGGAAACCAACCGCT	1075

RESIST 3

```

US-10-084-817-120
; Sequence 120, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Pion
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 120
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 3282941CB1
US-10-084-817-120

```

Alignment Scores:	
Pred. No.:	3.34e-54
Length:	2196

[illegible]

REC'D. T. A.

```

RESULT 4
US-10-240-965-83
; Sequence 83, Application US/10240965
; Publication No. US20030165924A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: SHOFFMAN, Dov
; APPLICANT: SOMOGYI, Roland
; APPLICANT: LAWN, Richard M.
; APPLICANT: SEILHAMER, Jeffrey J.
; APPLICANT: PORTER, Gordon J.
; APPLICANT: MIKITA, Thomas
; APPLICANT: TAL, Julie
; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION
; FILE REFERENCE: PA-0025 PCT
; CURRENT APPLICATION NUMBER: US/10/240,965
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: 60/195,106
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PERL Program

```



```

; SEQ ID NO 9069
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(482)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-9069

Alignment Scores:
Pred. No.: 1,04e-51 Length: 482
Score: 487.00 Matches: 95
Percent Similarity: 98.96% Conservative: 0
Best Local Similarity: 98.95% Mismatches: 1
Query Match: 53.11% Indels: 0
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-918-995-9069 (1-482)

Qy 84 AlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLys 103
Db 33 GCNTAGCCGCGAGAGCGGCCCGCATGGAGTGTGAGCAAGCCCTCTCCAGACTCAAG 92
Qy 104 ThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThrLeuArgLeu 123
Db 93 ACCACCCTCCCTGGGTGGTGGCCGCCACACCAAGCTCTCCAAAGTGGACACGCTCAGCGTG 152
Qy 124 AlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGly 143
Db 153 GCGTCCAGCTAGATGCCCACTTGAGGCAGATCTGCTAACGACAAATACGAGACCGG 212
Qy 144 TyrIleHisProValAsnLeuThrTrpProPheMetValalaGlyLysProGluSerAsp 163
Db 213 TACATTACCCCGTCAACCTGACGTGGCCCTTTATGTGGCCGGGAAACCCGAGAGTGAC 272
Qy 164 LeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 273 CTGAAGAAGTGGTACCGCGAGCCGCTTATGTGGACCAACCACCGCGTCC 320

RESULT 7
US-10-755-889-365
; Sequence 365, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 365
; LENGTH: 1939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-365

Alignment Scores:
Pred. No.: 3,05e-50 Length: 1939
Score: 482.00 Matches: 112
Percent Similarity: 62.26% Conservative: 20
Best Local Similarity: 52.83% Mismatches: 40
Query Match: 52.56% Indels: 41
DB: 17 Gaps: 5

US-09-701-674A-23 (1-179) x US-10-755-889-365 (1-1939)

```

Qy	1	MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluMetLeuGluCys	20
Db	213	ATGTCACGGGCTCGGTGAGTGATCCGGAGGAG	254
Qy	21	AspGlyLeuLys-	25
Db	255	CGGGGGCTGCACGGGAGTACCGGCTCCGAGAGCGCCCTTCGCGCGGTA	314
Qy	26	AspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu	40
Db	315	GAGCGCAGCTACGCTCCGCCAGTGACAACTCGTCGCGAGAGAGGAGGACCCCGACGGC	374
Qy	41	-	48
Db	375	GAGGAGGAGCGTGGCTCTGGGCACACCGCGCAGCGCGGAGGCTGCAAGAGGAGCGG	434
Qy	49	ProGln---LysGlyArgGlyGlyLeuGlyLysArgLysAlaProThrLysLysSer	67
Db	435	CCCCGTGTGGCTGGCGGGCGGCGCAGGTGTGAGTCGCGGGCGTGTGTGCAAGAGCC	494
Qy	68	---ProLeuSerGlyValSerGlnGlnGlyLysGlnValGlnArgAsnAlaAlaAsnAla	86
Db	495	CTCCCGGCCAAGGCTCAGCCGCGAGTGCAGGAGTCCGAGCGGAAACGGCGCCACAGCC	554
Qy	87	ArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeu	106
Db	555	CGTAGCGTGCCTCGGATCGCGTGCAGGCAAGACCTTCTCCAGGCTCAAGACACGAGCTG	614
Qy	107	ProTrpValProAspThrLysLeuSerLysLysAlaPheSerArgLeuLysThrThrLeu	126
Db	615	CCCTGGGTGCCCCCGACACTAAAGCTCTCCAAAGCTGGACACGCTCGGCTGGGCTCCAGT	674
Qy	127	TyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsnGlyTyrIleHis	146
Db	675	TACATCGCTCACTCGGCGAGCTGTCAGAGGAGCCGCTATGAGAACGCGTACGTGCAC	734
Qy	147	ProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAspLeuLysGlu	166
Db	735	CCAGTGTAACCTCAGATGGCCATTCGTGGTCTCTCGGGAAGACCGGCTCTGC-ACCAAAGAA	793
Qy	167	ValValThrAlaSerArgLeuCysGlyThrThrAla	178
Db	794	GTTCCGCGAGCCACAGACTATGTGGAAACACCGCT	829
RESULT 8			
US-10-305-720-1028			
; Sequence 1028, Application US10305720			
; Publication No. US20040010136A1			
; GENERAL INFORMATION:			
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.			
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene			
; FILE REFERENCE: PA-0002-1 CON			
; CURRENT APPLICATION NUMBER: US/10/305,720			
; CURRENT FILING DATE: 2002-11-26			
; PRIOR APPLICATION NUMBER: 09/016,434			
; PRIOR FILING DATE: 1998-01-30			
; NUMBER OF SEQ ID NOS: 1490			
; SOFTWARE: PERL Program			
; SEQ ID NO 1028			
; LENGTH: 240			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; OTHER INFORMATION: Incyte ID No. US20040010136A1 954226			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (1) ... (240)			
; OTHER INFORMATION: a, t, c, g, or other			
US-10-305-720-1028			
Alignment Scores:			
Pred. No.: 3,13e-35 Length: 240			

Score:	353.00	Matches:	76
Percent Similarity:	93.83%	Conservative:	3
Best Local Similarity:	93.83%	Mismatches:	3
Query Match:	38.50%	Indels:	2
DB:	16	Gaps:	0

US-09-701-674A-23 (1-179) x US-10-305-720-1028 (1-240)

Qy	61	LysAlaProThrLysLysSerProLeuSerGlyValSerGln-GluGlyLysGlnValGd	80
Db	1	AAGGGCCCCACCAGAAGAGCCCCCTGAGCGGGCTAGCCAGGAGGGGAGCAGGTCCA	60
Qy	80	nargAsnAlaAlaAsnAlaArgGluArg-AlaArgMetArgValLeuSerLysAlaPheS	100
Db	61	GGCGAACGNCGGCCACGGCGAGAGCGGGGCCCATGCGAGTGTGAGCAAGCGCTCT	120
Qy	100	erArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspT	120
Db	121	CCAGACTCAAGACACCCCTGGCCCTGGGTGCCCCCGGCACACCAAGCTCTCCAGCTGGACA	180
Qy	120	hrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLys	139
Db	181	CGCTCAGGCTGGCGTCCAGNACATNGCCACTTGAGCGAGATCTCGGTACGACAAA	239

## RESULT 9

US-09-954-456-829  
; Sequence 829, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer Agents;  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456

; APPLICANT: Young, Paul  
 ; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

Alignment Scores:		
Pred. No.:	3.54e-30	697
Score:	318.00	61
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	34.68%	Indels: 0
DB:	9	Gaps: 0

US-09-701-674A-23 (1-179) x US-09-954-456-829 (1-697)

Qy	119	AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp	138
Db	3	GACAGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGGCAGATCTCGGTAAACGAC	62
Qy	139	LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly	158
Db	63	AAATACGAGAACGGGTACATTACCCCGGTCAACCTGACGTGGCGCCTTTATGTGGCGCGG	122
Qy	159	LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla	178
Db	123	AAACCCGAGAGTGACCTGAAGAAGAGTGGTACCCGAGCCGCTTATGTGGNACCACCGG	182
Qy	179	Ser	179
Db	183	TCC	185

## RESULTS

US-09-954-456-1245  
; Sequence 1245, Application US/09954456  
; Patent No. US20020115057A1  
GENERAL INFORMATION:

	Pred. No.: 3	3.54e-30	Length:	697
	Score:	318.00	Matches:	61
	Percent Similarity:	100.00%	Conservative:	0
	Best Local Similarity:	100.00%	Mismatches:	0
	Query Match:	34.68%	Indels:	0
	DB:	9	Gaps:	0
Alignment Scores:				

US-09-701-674A-23 (1-179) X US-09-954-456-1245 (1-697)

Qy 119 AspThrIeuArgIeuAlaSerSertyrIleAlaHisIeuArgGinIleLeuAlaAsnAsp 138



```
Db      3  GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGCGAGATCCTGGCTAACGAC 62
Qy      139  LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158
          |||
Db      63  AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCGTGGCCCTTTATGTGGCCGGG 122
          |||
Qy      159  LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178
          |||
Db      123  AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCAACCGCG 182
          |||
Qy      179  Ser 179
          |||
Db      183  TCC 185

RESULT 11
US-09-960-706-1028
; Sequence 1028, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1028
; LENGTH: 697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 W73859
; NAME/KEY: unsure
; LOCATION: (1)..(697)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-960-706-1028

Alignment Scores:
Pred. No.:      3,54e-30      Length:      697
Score:          318.00      Matches:      61
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    34.68%      Indels:      0
DB:             10      Gaps:      0

US-09-701-674A-23 (1-179) x US-09-960-706-1028 (1-697)

Qy      119  AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138
          |||
Db      3  GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGCGAGATCCTGGCTAACGAC 62
          |||
Qy      139  LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158
          |||
Db      63  AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCGTGGCCCTTTATGTGGCCGGG 122
          |||
Qy      159  LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178
          |||
Db      123  AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCAACCGCG 182
          |||
Qy      179  Ser 179
          |||
Db      183  TCC 185

RESULT 12
US-09-873-319-679
; Sequence 679, Application US/09873319A
; Publication No. US20030134324A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Munger, William E.
; APPLICANT: Kulkarni, Prakash
; APPLICANT: Getzenberg, Robert H.
; APPLICANT: Waga, Iwao
; APPLICANT: Yamamoto, Jun
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia Using Gene Expression Profiles
; FILE REFERENCE: 44921-5029-US
; CURRENT APPLICATION NUMBER: US/09/873,319A
; CURRENT FILING DATE: 2001-06-05
; EARLIER APPLICATION NUMBER: US 60/223,323
; EARLIER FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 755
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 679
; LENGTH: 697
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 W73859
; NAME/KEY: unsure
; LOCATION: (1)..(697)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-873-319-679

Alignment Scores:
Pred. No.:      3,54e-30      Length:      697
Score:          318.00      Matches:      61
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    34.68%      Indels:      0
DB:             10      Gaps:      0

US-09-701-674A-23 (1-179) x US-09-873-319-679 (1-697)

Qy      119  AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138
          |||
Db      3  GACACGCTCAGGCTGGCGTCCAGCTACATCGCCACTTGAGCGAGATCCTGGCTAACGAC 62
          |||
Qy      139  LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158
          |||
Db      63  AAATACGAGAACGGGTACATTACCCGGTCAACCTGAGCGTGGCCCTTTATGTGGCCGGG 122
          |||
Qy      159  LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178
          |||
Db      123  AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCGCTTATGTGGAACCAACCGCG 182
          |||
Qy      179  Ser 179
          |||
Db      183  TCC 185

RESULT 13
US-09-918-995-11181
; Sequence 11181, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11181
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: (1)...(493)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-11181

Alignment Scores:
Pred. No.: 4,9e-19 Length: 493
Score: 228.00 Matches: 42
Percent Similarity: 97.73% Conservative: 1
Best Local Similarity: 95.45% Mismatches: 1
Query Match: 24.86% Indels: 0
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-918-995-11181 (1-493)

Qy 136 AlaAsnAspLysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMet 155
Db 54 TCTAAGCACAATACGAGAACGGGTACATTACCCGGTCAACCTGATGGCCCTTTATG 113
Qy 156 ValAlaGlySerProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGly 175
Db 114 GTGGCCGGGAACCCGAGAGTGACCTGAAGAAGTGGTGACCGCGAGCGCTTATGTGA 173
Qy 176 ThrThrAlaSer 179
Db 174 ACCACCGCGTCC 185

RESULT 14
US-10-450-826-124
; Sequence 124, Application US/10450826
; Publication No. US20040101818A1
; GENERAL INFORMATION:
; APPLICANT: Ji, Darren
; APPLICANT: Axelrod, Douglas W.
; APPLICANT: Cook, Jonathan S.
; APPLICANT: Jaiswal, Neelam
; APPLICANT: Eistein, Richard
; APPLICANT: Houghton, Adam
; APPLICANT: Mertz, Lawrence
; TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiation
; FILE REFERENCE: 044921-5039-WO
; CURRENT APPLICATION NUMBER: US/10/450,826
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/255,882
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/285,691
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 124
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. S78079
US-10-450-826-124

Alignment Scores:
Pred. No.: 1,53e-12 Length: 1140
Score: 180.50 Matches: 46
Percent Similarity: 60.19% Conservative: 19
Best Local Similarity: 42.59% Mismatches: 32
Query Match: 19.68% Indels: 11
DB: 17 Gaps: 3

US-09-701-674A-23 (1-179) x US-10-450-826-124 (1-1140)

Qy 39 GluGluSerSerAsnCysGluAsnGlySerProGlnLys-----Gly 52
Db 243 GAGGACCGGGAGGAGAGCTCGGCTCCGACGAGAACCCCTGCGTGTGATGCTGCG 302
Qy 53 ArgGlyGlyLeu---GlyLysArgGlyAlaProThrLysLysSerProLeuSerGly 71
Db 303 CGCTGTGGCTCCAGGGCCCGGGGGGCGGAGGAGGAGGCGGGTACGCTGCTGCT 362

; LOCATION: (1)...(493)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-11181

Alignment Scores:
Pred. No.: 4,9e-19 Length: 493
Score: 228.00 Matches: 42
Percent Similarity: 97.73% Conservative: 1
Best Local Similarity: 95.45% Mismatches: 1
Query Match: 24.86% Indels: 0
DB: 10 Gaps: 0

US-09-701-674A-23 (1-179) x US-09-918-995-11181 (1-493)

Qy 136 AlaAsnAspLysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMet 155
Db 54 TCTAAGCACAATACGAGAACGGGTACATTACCCGGTCAACCTGATGGCCCTTTATG 113
Qy 156 ValAlaGlySerProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGly 175
Db 114 GTGGCCGGGAACCCGAGAGTGACCTGAAGAAGTGGTGACCGCGAGCGCTTATGTGA 173
Qy 176 ThrThrAlaSer 179
Db 174 ACCACCGCGTCC 185

RESULT 14
US-10-450-826-124
; Sequence 124, Application US/10450826
; Publication No. US20040101818A1
; GENERAL INFORMATION:
; APPLICANT: Ji, Darren
; APPLICANT: Axelrod, Douglas W.
; APPLICANT: Cook, Jonathan S.
; APPLICANT: Jaiswal, Neelam
; APPLICANT: Eistein, Richard
; APPLICANT: Houghton, Adam
; APPLICANT: Mertz, Lawrence
; TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiation
; FILE REFERENCE: 044921-5039-WO
; CURRENT APPLICATION NUMBER: US/10/450,826
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/255,882
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/285,691
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 124
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. S78079
US-10-450-826-124

Alignment Scores:
Pred. No.: 1,53e-12 Length: 1140
Score: 180.50 Matches: 46
Percent Similarity: 60.19% Conservative: 19
Best Local Similarity: 42.59% Mismatches: 32
Query Match: 19.68% Indels: 11
DB: 17 Gaps: 3

US-09-701-674A-23 (1-179) x US-10-450-826-124 (1-1140)

Qy 39 GluGluSerSerAsnCysGluAsnGlySerProGlnLys-----Gly 52
Db 243 GAGGACCGGGAGGAGAGCTCGGCTCCGACGAGAACCCCTGCGTGTGATGCTGCG 302
Qy 53 ArgGlyGlyLeu---GlyLysArgGlyAlaProThrLysLysSerProLeuSerGly 71
Db 303 CGCTGTGGCTCCAGGGCCCGGGGGGCGGAGGAGGAGGCGGGTACGCTGCTGCT 362

```

us-09-701-674a-23.rnpb

Fri Dec 17 09:25:28 2004

Job time : 491 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 21:48:16 ; Search time 2870 Seconds  
(without alignments)  
2272.722 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MTGSLSDVEDIQEVMLEEC.....PESDLKEVTVASRLCGTTAS 179

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USPTO.spool\_h/US09701674/runat\_15122004\_100441\_27517/app\_query.fasta\_1.327  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXI=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 @CNG 1 1 3437 @runat\_15122004\_100441\_27517 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_hic:  
4: gb\_est3:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_gss1:  
9: gb\_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	917	100.0	782	CF128484	UI-HF-ETO
2	917	100.0	796	EX107489	BM544256
3	917	100.0	1044	AL550251	AL550251
4	917	100.0	1221	CR594669	full-length
5	917	100.0	1250	CR606057	full-length
6	917	100.0	1252	CR616308	full-length
7	912	99.5	1038	BM544256	AGENCOURT
8	911	99.3	1024	AL575799	AL575799
9	910	99.2	1099	AL552119	AL552119

10	909	99.1	800	6	CF127519	UI-HF-ETO
11	901	98.3	827	5	EX926295	EX926295
12	901	98.3	1098	4	BM923344	AGENCOURT
13	900	98.1	1056	5	EX337699	EX337699
14	890	97.1	1085	1	AL574170	AL574170
15	887	96.7	1012	6	BY710826	BY710826
16	887	96.7	1238	3	AK011575	Mus muscu
17	883	96.3	996	5	EX360244	EX360244
18	876	95.5	778	7	CO564290	AGENCOURT
19	876	95.5	965	1	AL552380	AL552380
20	874	95.3	794	6	CF127034	UI-HF-ETO
21	870	94.9	726	5	EX920031	EX920031
22	867	94.5	1169	1	AL552084	AL552084
23	859	93.7	647	7	CF359503	813267 MA
24	855	93.2	712	5	EX922447	EX922447
25	852	92.9	926	4	BI769090	603057070
26	851	92.8	563	1	AV605533	AV605533
27	847	92.4	543	2	BF198267	BF198267
28	845.5	92.2	1065	5	BQ072411	AGENCOURT
29	839	91.5	740	5	BQ290232	603577325
30	839	91.5	879	5	BU459258	603366192
31	836	91.2	818	4	BI767574	603063488
32	827	90.2	597	5	BQ358809	603477646
33	821	89.5	746	4	BI770998	603055341
34	820	89.4	785	6	CF127896	UI-HF-ETO
35	810	88.3	715	7	CK835890	4060455 B
36	810	88.3	718	7	CN409630	170004245
37	800	87.2	757	6	CF126927	UI-HF-ETO
38	799	87.1	675	5	EX919517	EX919517
39	799	87.1	792	6	CF270487	AGENCOURT
40	798	87.0	645	6	CF128246	UI-HF-ETO
41	797	86.9	589	6	CB417689	590416 MA
42	797	86.9	744	6	CF128914	UI-HF-ETO
43	783.5	85.4	826	4	BI768711	603057274
44	781	85.2	717	6	CF126467	UI-HF-ETO
45	781	85.2	837	4	BI770596	603060402

#### ALIGNMENTS

CF128484 782 bp mRNA linear EST 05-AUG-2003  
UI-HF-ETO-awj-c-15-0-UI.r1 NIH MGC\_214 Homo sapiens CDNA clone  
IMAGE:30557534 5', mRNA sequence.

CF128484  
CF128484.1 GI:33207776

EST.  
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 782)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

PUBMED

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Clone Sequencing by: Dr. M. Bento Soares, University of Iowa

http://genome.uiowa.edu/distribution/humanfl.html

The following repetitive elements were found in this cDNA



QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 375 AAGCGCCCAACCAAGAGAGCCCTTGCAGCGGGTTCAGCGAGGAGGAGGAGTCCAG 434  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 435 CGAAACGGCCCAACCGCGGAGAGCGGGCCGATCGAGTCTGAGCAAGCCCTTCTCC 494  
 QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120  
 Db 495 AGACTCAAGACCAACCTGCTGGTGGTGGCCCGACCAAGCTCTCCAAAGTGGACAG 554  
 QY 121 LeuArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnlleAlaAlaSerLysTyrl 140  
 Db 555 CTCAGCTGCGTCCAGCTACATCGCCCACTTGGAGCAGATCTTGGCTTAAACACAAATAC 614  
 QY 141 GluAsnGlyTyrlleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 615 GAGAACGGGTACATTCAACCGGTCAACCTGAGTGGCCCTTTATGGTGGCCGGAAACCC 674  
 QY 161 GluSerAspLeuLysGlnValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 675 GAGAGTGACCTGAAGAAGTGTGACCGGAGCGCGCTTATGTGGAACCAACCGCGTCC 731

RESULT 3  
 AL550251 1044 bp mRNA linear EST 25-MAR-2004  
 LOCUS  
 DEFINITION  
 AL550251 Homo sapiens PRIMATE COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0D1039YJ05 5-PRIME, mRNA sequence.

ACCESSION  
 AL550251  
 VERSION  
 AL550251.3 GI:45750628  
 KEYWORDS  
 EST.

SOURCE  
 Homo sapiens (human)

ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

On Feb 15, 2001 this sequence version replaced gi:31272068.

Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f

For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?s=CS0D1039CE03QP1&c=3812.f.

FEATURES  
 source

Location/Qualifiers

1..1044  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1039YJ05"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f"

ORIGIN

Alignment Scores:  
 Pred. No.: 1,98e-86 Length: 1044  
 Score: 917.00 Matches: 179  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0

DB: 1 Gaps: 0  
 US-09-701-674A-23 (1-179) x AL550251 (1-1044)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 278 ATGTCACCGGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGGAATGT 337  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 338 GACGGGTGAAATGGATTGGAACAGGAATTTGTGACTTCCACAGAGACCCAGGAG 397  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 Db 398 AGCTCCAACTCGGAGATGGGTCTCCCCAGAAAGGGCCGCGCGCTGGGCAAGAGAGG 457  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 458 AAGCGGCCCAACCAAGAGAGCCCTTGCAGCGGGTTCAGCCAGAGGAGGAGCGTCCAG 517  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 518 CGCAACCGCCCAACCGCGGAGAGCGGGCCGATCGAGTCTGAGCAAGCCCTTCTCC 577  
 QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLysLeuSerLysLeuAspThr 120  
 Db 578 AGACTCAAGACCAACCTTGGTGGTGGCCCGACCAAGCTCTCCAAAGTGGAGACG 637  
 QY 121 LeuArgLeuAlaSerSerTyrlleAlaHisLeuArgGlnlleAlaAlaSerLysTyrl 140  
 Db 638 CTCAGGCTGGCTCCAGCTACATCGCCCACTTGGAGCAGATCTTGGCTTAAACACAAATAC 697  
 QY 141 GluAsnGlyTyrlleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 698 GAGAACCGGTACATTCAACCGGTCAACCTGAGTGGCCCTTTATGGTGGCCGGAAACCC 757  
 QY 161 GluSerAspLeuLysGlnValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 758 GAGAGTGACCTGAAGAAGTGTGACCGGAGCGCGCTTATGTGGAACCAACCGCGTCC 814

RESULT 4

CR594669

LOCUS

DEFINITION

full-length cDNA clone CS0D1059YH17 of Placenta Cot 25-normalized

of Homo sapiens (human).

ACCESSION

CR594669

VERSION

CR594669.1 GI:50475476

KEYWORDS

HTC; CNSUT\_cDNA.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1221)

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

REMARK

Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Faraday Avenue

2 (bases 1 to 1221)

Genoscope.

REFERENCE

AUTHORS

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

Web : www.genoscope.cns.fr)

COMMENT

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and EcoR

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen.

FEATURES

Location/Qualifiers

1..1221

/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039YJ05"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="PCMVSPORT_6"

ORIGIN
Alignment Scores:      2,45e-86      Length:      1221
Pred. No.:            917.00      Matches:      179
Score:                100.00%      Conservative:  0
Percent Similarity:   100.00%      Mismatches:   0
Best Local Similarity: 100.00%      Indels:       0
Query Match:         100.00%      Gaps:        3
DB:                  3

US-09-701-674A-23 (1-179) x CR594669 (1-1221)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
DB 260 ATGTCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTGGAATGT 319
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
DB 320 GACGGGTTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCACGAGACGACCGAGGAG 379
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
DB 380 AGCTCCAACTCGAGATGGGCTCTCCCAAGAGGGCGCGGCTGGGCAAGAGGAGG 439
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
DB 440 AAGGCGCCCAACCAAGAGAGCGCCCTGAGCGGGGTGAGCAGGAGGAGGAGGAGTCCAG 499
QY 81 ArgAsnAlaAlaSerThrLeuArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
DB 500 CGCAACGCCGCCCAACCGCGAGAGCGGGCCCGCATGCGAGTGGTCTGAGCAGGCTTCTCC 559
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
DB 560 AGACTCAAGACCACTGCGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 619
QY 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140
DB 620 CTGAGGTGGCGTCCAGTACATCGCCCACTTGAAGCAGATCTCTGAGTAAACGCAATAC 679
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
DB 680 GAGACGGGTACATTCACCGCGTCAACCTGACGTGGCCCTTTATGTTGGCCCGGGAACCC 739
QY 161 GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
DB 740 GAGAGTGACCTGAAGAAGTGGTGGCGGAGCGGCTTATGTGGAACCAACCGCGTCC 796

RESULT 5
CR606057      1250 bp      mRNA      linear      HTC 21-JUL-2004
LOCUS      full-length cDNA clone CS0D1039YJ05 of Placenta Cot 25-normalized
DEFINITION      of Homo sapiens (human).
ACCESSION      CR606057
VERSION      CR606057.1 GI:50486864
KEYWORDS      HTC; CDS; cDNA.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1250)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1250)

REFERENCE
AUTHORS      CR616308      1252 bp      mRNA      linear      HTC 21-JUL-2004
TITLE      full-length cDNA clone CS0D1070YK16 of Placenta Cot 25-normalized
JOURNAL      of Homo sapiens (human).
REMARK      CR616308
ACCESSION      CR616308.1 GI:50497115
VERSION

```

```

Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source      Location/Qualifiers
1..1250
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039YJ05"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="PCMVSPORT_6"

ORIGIN
Alignment Scores:      2,53e-86      Length:      1250
Pred. No.:            917.00      Matches:      179
Score:                100.00%      Conservative:  0
Percent Similarity:   100.00%      Mismatches:   0
Best Local Similarity: 100.00%      Indels:       0
Query Match:         100.00%      Gaps:        3
DB:                  3

US-09-701-674A-23 (1-179) x CR606057 (1-1250)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
DB 278 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAAGAGTGGAGATGTGGAATGT 337
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
DB 338 GACGGGTTGAAATGGAATTCGAACAAGGAATTTGTGACTTCCAAACGAGACGACCGAGGAG 397
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
DB 398 AGTCTCAACTGCGAGATGGGTCTCCCGCAGAGGGCGCGGCGCTGGGCAAGAGGAGG 457
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
DB 458 AAGCGGCCCAACCAAGAGAGACCCCTGAGCGGGGTGAGCAGGAGGAGGAGTCCAG 517
QY 81 ArgAsnAlaAlaSerThrLeuArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
DB 518 CGCAACGCCGCCCAACCGCGAGAGCGGGCCCGCATGCGAGTGGTGGAGCAAGGCTTCTCC 577
QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120
DB 578 AGACTCAAGACCACTGCGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637
QY 121 LeuArgLeuAlaSerSerThrLeuAlaHisLeuArgGlnLeuAlaAsnAspLysTyr 140
DB 638 CTCAGGTGGCGTCCAGTACATCGCCCACTTGAAGCAGATCTCTGAGTAAACGCAATAC 697
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
DB 698 GAGAACGGGTACATTCACCGCGTCAACCTGAGTGGCCCTTTATGTTGGCCCGGGAACCC 757
QY 161 GluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
DB 758 GAGAGTGACCTGAAGAAGTGGTGGCGGAGCGGCTTATGTGGAACCAACCGCGTCC 814

RESULT 6
CR616308      1252 bp      mRNA      linear      HTC 21-JUL-2004
LOCUS      full-length cDNA clone CS0D1070YK16 of Placenta Cot 25-normalized
DEFINITION      of Homo sapiens (human).
ACCESSION      CR616308
VERSION      CR616308.1 GI:50497115

```



```

KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 1252)
REMARK Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1252)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1252
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS01070YK16"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.: 2 54e-86 Length: 1252
Score: 917.00 Matches: 179
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-701-674A-23 (1-179) x CH616308 (1-1252)
QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
272 ARGTCACCGCGTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTTGGAATGT 331
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
332 GACGGTTGAAATGGATTGCAACAGGAATTTGTGACTTCCACGAGACCCAGGAG 391
QY 41 SerSerAsnGlyGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60
392 AGCTCCAACTGGCAGAAATGGGTCTCCCAAGAGGGCGCGCGGCTGGCAAGAGGAGG 451
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
452 AAGGCCCCACCACCAAGAGAGCCCCCTGACGGGGTTCAGCCAGAGGGGGAAGCAGTCCAG 511
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
512 CGCAACGCCGCCAACCGCGAGAGCGGCCCGCATGGCAGTGTGAGCAAGGCCTTCTCC 571
QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120
572 AGACTCAAGACCACTCGCCCTGGGTGGTGGCCCGCCGACCAAGACTCTCCAAGCTGGACAG 631
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
632 CTCAGGTGGCGTCCAGTACTATCCCACTTGGAGGAGATCTCTGGCTTACCAATATAC 691
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160

```

```

Db 692 GAGAACGGGTACATTCACTCCCGGTCAACCTGACGTGGCGCCCTTTATGGTGGCGGAAACCC 751
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
752 GAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCTTATGTGGAACCCAGCGCTCC 808
RESULT 7
BM544256
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1038)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12357 Row: m Column: 09
High quality sequence stop: 774.
FEATURES
Location/Qualifiers
1..1038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5587760"
/lab_host="DH10B"
/clone_lib="NIH_MGC_125"
/notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site1: EcoRV (destroyed); Site2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."
ORIGIN
Alignment Scores:
Pred. No.: 6 65e-86 Length: 1038
Score: 912.00 Matches: 178
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.45% Indels: 0
DB: 4 Gaps: 0
US-09-701-674A-23 (1-179) x BM544256 (1-1038)
QY 2 SerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAsp 21
82 TCCACCGCGTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTTGGAATGTGAC 141
QY 22 GlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSer 41
142 GGGTTGAAATGGATTTCGAACAGGAATTTGTGACTTCCACGAGACCCAGGAGAGC 201
QY 42 SerAsnGlyGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArgLys 61
202 TCCAACTCGAGAAATGGGTCTCCCCAGAGGGCGCGGCTCTGGCAAGAGGAGGAG 261

```



```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI059YH1.7"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Alignment Scores:
  Length: 1178-85
  Score: 910.00
  Percent Similarity: 99.44%
  Best Local Similarity: 99.44%
  Query Match: 99.24%
  DB: 1
  Gaps: 0
  Indels: 0
  Mismatches: 0
  Conservatives: 178
  Matches: 1099

US-09-701-674A-23 (1-179) x AL552119 (1-1099)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 260 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAGAGGTGGAGATGTTGAATGT 319
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 320 GACGGGTGAAATGGAATCGAACAGGAATTTGTGACTTCCACGAGACCCGAGGAG 379
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
Db 380 AGCTCCAAATGGAGAAATGGTCTCCCAAGAGGGCGCGCGCTGGCGCAAGAGGAGG 439
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 440 AAGCGCCCAACAGAGAGCCCTCAGCGGGTCAGCCAGAGGGGAGGAGCAGGTCCAG 499
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 500 CGCAAGCGCCCAACAGCGGAGCGGGCCCGCATGCGAGTCTCAGCAAGCGCTTCCTCC 559
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120
Db 560 AGACTCAAGACCACTCGCTCGCTGGTGGCCCGCCGACCAAGCTTCCAGCTGGACAG 619
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db 620 CTCAGCTGGCGTCCAGCTACATCGCCCACTTGAAGGACATCCTGGCTAAACGACAAATAC 679
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db 680 GAGACGGGTATCATTCACCGCTCAACCTGACGTGGCCCTTTATGGTGGCGGGAACCC 739
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuLysGlyThrThrAlaSer 179
Db 740 GAGAGTKACCTGAAGAGTGTGTACCGCGAGCCGCTTATGTGGAACCAACCGCGTCC 796

```

## RESULT 10

```

CF127519
LOCUS CF127519
DEFINITION UI-HF-ET0-awh-f-14-0-UI.r1 NIH MGC_214 Homo sapiens cDNA clone
IMAGE:30554533 5', mRNA sequence.
CF127519
EST. CF127519.1 GI:33205841
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 800)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

```

```

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT

```

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9585  
Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Cloning Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/humanfl.html  
The following repetitive elements were found in this cDNA  
sequence: 172-253, >(GA)n#simple\_repeat (matched complement)  
Seq primer: pYX-5.

## FEATURES

Location/Qualifiers

## Source

```

1..800
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30554533"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH MGC_214"
/notes="Organ: Lung; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATAAGGCCA. Tissue was provided by Mary Hendrix."

```

## ORIGIN

```

Alignment Scores:
  Length: 800
  Score: 909.00
  Percent Similarity: 100.00%
  Best Local Similarity: 100.00%
  Query Match: 99.13%
  DB: 6
  Gaps: 0

US-09-701-674A-23 (1-179) x CF127519 (1-800)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 258 ATGTCCACCGCTCCCTCAGCATGTGGAGGACCTTCAGAGGTGGAGATGTTGAATGT 317
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 318 GACGGGTGAAATGGAATCGAACAGGAATTTGTGACTTCCACGAGACCCGAGGAG 377
QY 41 SerSerAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60
Db 378 AGCTCCAAATGGAGAAATGGTCTCCCAAGAGGGCGCGCGCTGGCGCAAGAGGAGG 437
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
Db 438 AAGCGCCCAACAGAGAGCCCTCAGCGGGTCAGCGGGGTGAGCGAGGAGGAGGAGG 497
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 498 CGCAAGCGCCCAACAGCGGAGAGCGGGCCCGCATGCGAGTCTGAGCAAGCGCTTCCTCC 557
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120

```



stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

## ORIGIN

## Alignment Scores:

Pred. No.: 1,05e-84 Length: 1098  
Score: 901.00 Matches: 178  
Percent Similarity: 99.44% Conservativeness: 1  
Best Local Similarity: 98.89% Mismatches: 0  
Query Match: 98.26% Indels: 1  
DB: 4 Gaps: 0

US-09-701-674A-23 (1-179) x BM923344 (1-1098)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
Db 268 ATGTCACCGCGCTCCCTCAGCGATGTGGAGACCTTCAAGAGGTGGAGATGTTGAATGT 327  
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
Db 328 GACGGGTGAATGATTCGACACAGGAATTTGTGACTTCCACGAGACCGAGGAG 387  
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyLeuLysArgArg 60  
Db 388 AGCTCCAACTGCGAGAAATGGTGTCCCCAGAGGGCGCGCGGCTGGCGAAGAGGAGG 447  
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluLysGlnValGln 80  
Db 448 AAGCGCCGCCACCAAGAGAGCCCTCAGCGGGTCCGACAGAGGGAGCAGGTCCAG 507  
QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
Db 508 CGCACCGCGCCAAACGCGAGAGCGCGCGCATCGAGTGTCTGAGCAAGCCCTTCTCC 567  
QY 101 ArgLeuLysThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThr 120  
Db 568 AGACTCAAGACCACTCGCCCTGGGTGGTCCCGCCGACACCAAGCTCTCCAGCTGACACG 627  
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
Db 628 CTCAGCTGGGTCCAGCTACATCGCCCACTTGAGGCAGATCTTGCTAACGACAAATAC 687  
QY 141 GluAsnGlyTyrIleHisProValLeuThrTrpProPheMetValAlaGlyLysPr 160  
Db 688 GAGAACGGGTACATTCACCGGTCAACCTGAGCTGGCCCTTTATGTTGGCGGGAACC 747  
QY 160 oGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
Db 748 CGAGAGTGACCTTGAAAGAAATGTTGACCGCGAGCGCCCTTATGTGAAACCCCGCTCC 805

## RESULT 13

## BX337699

## LOCUS

## DEFINITION

## BX337699 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

## BX337699

## BX337699

## EST.

## KEYWORDS

## SOURCE

## Homo sapiens

## ORGANISM

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## 1 (bases 1 to 1056)

## Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

## Full-length cDNA libraries and normalization

## Unpublished (2001)

## On May 2, 2003 this sequence version replaced gi:30345636.

## COMMENT

## Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1048AC08QPl&c=3812.f.

Location/Qualifiers

1..1056

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1048VE15"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.

Library was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1048AC08QPl&c=3812.f.

Location/Qualifiers

1..1056

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1048VE15"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.

Library was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1048AC08QPl&c=3812.f.

Location/Qualifiers

1..1056

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1048VE15"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.

Library was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1048AC08QPl&c=3812.f.

Location/Qualifiers

1..1056

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1048VE15"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.

Library was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1048AC08QPl&c=3812.f.

Location/Qualifiers

1..1056

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D1048VE15"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

LOCUS AL574170 1085 bp mRNA linear EST 06-APR-2004  
 DEFINITION Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
 clone CS0D1039YJ05 3-PRIME, mRNA sequence.  
 ACCESSION AL574170  
 VERSION AL574170.3 GI:46247266  
 EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1085)  
 AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 16, 2001 this sequence version replaced gi:31295505.

Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?s=CS0D1039CE03NP1&c=3812.f.

FEATURES  
 source  
 1..1085  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1039YJ05"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,51e-83 Length: 1085  
 Score: 890.00 Matches: 178  
 Percent Similarity: 99.44% Conservative: 0  
 Best Local Similarity: 99.44% Mismatches: 1  
 Query Match: 97.06% Indels: 1  
 DB: 1 Gaps: 0

US-09-701-674A-23 (1-179) x AL574170 (1-1085)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20  
 Db 969 ATGTCACCGGCTCCCTCAGCATGTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 910  
 QY 21 AspGlyLeuLeuMetAspSerAnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 909 GACGGGTGTAATGATTCGAACAGGAATTTGTGACTTCCACGAGACCGGAGGAG 850  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArgArg 60  
 Db 849 AGCTCCAACTCGAGAATGGTCTCCAGAGAGGCGCGCGGCTGGCCAGAGGAGG 790  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerClnGluClyLysGlnValGln 80  
 Db 789 AAGGCGCCCAACAGAGAGCCCTTGGCGGGGTTCAGCGAGGGAGGAGGAGTCCAG 730  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluAlaArgMetArgValLeuSerLysAlaPheSer 100  
 Db 729 CGCAACGCGCCCAACGCGGAGAGCGGCGCGGCTCGAGTGTGAGCAAGGCTTCTCC 670  
 QY 101 ArgLeuLysThrThrLeuProThrValProProAspThrLysLeuSerLysLeuAspThr 120

Db 669 AGACTCAAGACCAACCCCTGCCCTGGTG-CCCCCGGACACCAAGCTCTCCAGCTGGACAG 611  
 QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 610 CTCAGGCTGGCGTCCAGCTACATCGCCCACTTGGAGCAGATCTGGCTAACCAATAC 551  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 Db 550 GAGACGGGTACATTCAACCGGTCAACCTGACGTGGCCCTTTATGGTGGCCGGAACCC 491  
 QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 490 GAGAGTGACCTGAAGAAGTGGTGACCGGAGCGCTTATGGGAACCAACCCGCTCC 434

RESULT 15  
 LOCUS BY710826 1012 bp mRNA linear EST 16-DEC-2002  
 DEFINITION BY710826 RIKEN full-length enriched, 10 days embryo Mus musculus  
 CDNA clone 2610027010 5', mRNA sequence.

ACCESSION BY710826  
 VERSION BY710826.1 GI:27122059  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1012)  
 AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,

Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, I.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,  
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,  
 Batailov, S., Beisel, K. W., Blake, J. A., Bradt, D., Bruscia, V.,  
 Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, I. A.,  
 Fietch, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Gustinch, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,  
 Kawai, H., Kawasawa, Y., Kedierski, R. M., King, B. L., Konagaya, A.,  
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,  
 Maltais, B., Marchionni, L., McKenzie, L., Mik, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,  
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,  
 Sultan, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,  
 Verardo, R., Wagner, L. G., Wainwright, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inctani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
 Rogers, J., Birney, B. and Hayashizaki, Y.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 22354583  
 PUBMED 12466851

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
 Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
 Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
 Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
 Nemura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,  
 Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,

Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
 further details.

## FEATURES

Location/Qualifiers

Source

```

1..1012
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2610027010"
/sex="mixed"
/dev_stage="10 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days embryo"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTITTTTTTWN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCTCGATTAAATTAATCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I"
```

## ORIGIN

```

Alignment Scores:
Pred. No.: 2,84e-83 Length: 1012
Score: 887.00 Matches: 172
Percent Similarity: 98.32% Conservative: 4
Best Local Similarity: 96.09% Mismatches: 3
Query Match: 96.73% Indels: 0
DB: 6 Gaps: 0
```

US-09-701-674A-23 (1-179) x BY10826 (1-1012)

```

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCys 20
Db 241 ATGTCACCTGGCTCCCTCAGCGATGTAGAGACCTTCAGAGGTTGGAGATGCTGGACTGT 300
QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40
Db 301 GACTCCCTGAAGTGGACTCCACACAGGAGTTTGGAACTTCCAACGAGACCCGAGGAG 360
QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysArg 60
Db 361 GGCTCCACATCGCAGAACGGGCTCTCCACAGAGGGTCCGGTGGCTAGGCAAGAGAGG 420
QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80
```

```

Db 421 AAGCGGCCCTACTAAGAAAAGCCCGCTCAGCGGGGTACGCCAGGAGGCGCAAGCGTCCAG 480
QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100
Db 481 CGCAACGCGGCCCAATGCTCGTAGCGGGCCCGCATGCGGGTGTCTGAGCAAGGCGCTTCTCC 540
QY 101 ArgLeuLysThrThrLeuProTyrValProProAspThrLysLysLeuSerLysLeuAspThr 120
Db 541 AGGCTCAAGACCCACCCCTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
QY 121 LeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140
Db 601 CTCAGGCTGGCGTCCAGCTACATCGCTCACTTAAGGCAGATCCTGGCCAAACGACAAGTAC 660
QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160
Db 661 GAGAACCGTTTACATTCACCCAGTCAACCTGACGTGGGCCCTTTATGGTGGCGCGCAACCA 720
QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179
Db 721 GAGATGACCTGAAGGAAGTGGTGACCGCCCAACCGCTTGTGTGGAATACAGCATCC 777
```

Search completed: December 17, 2004, 00:03:53  
 Job time : 2876 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 17:03:57 ; Search time 4403 Seconds  
(without alignments)

10527.218 Million cell updates/sec

Title: US-09-701-674A-54

Perfect score: 1272

Sequence: 1 tctacggccacgactctggg.....attgtataaaaaaaaaaaaaa 1272

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_gss1:\*

9: gb\_gss2:\*

Pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1248	98.1	1252	3	CR616308 full-leng
2	1240	97.5	1250	3	CR606057 full-leng
3	1221	96.0	1221	3	CR594669 full-leng
C 4	1003.4	78.9	1085	1	AL574170 AL574170
5	995.8	78.3	1099	1	AL552119 AL552119
C 6	974	76.6	1024	1	AL575799 AL575799
7	963.8	75.8	1044	1	AL550251 AL550251
C 8	959.8	75.5	1169	1	AL552084 AL552084
9	939	73.8	1169	1	AL552380 AL552380
10	931	73.2	1056	5	BM544256 AGENCOURT
11	878.8	69.1	1038	4	BM544256 AGENCOURT
12	824	64.8	1098	4	BM923344 AGENCOURT
13	795	62.5	796	5	BM923344 AGENCOURT
14	788	61.9	996	5	BM923344 AGENCOURT
15	787	61.9	800	6	BM923344 AGENCOURT
16	770.4	60.6	794	6	CF127519 UI-HF-ETO
17	767.4	60.3	782	6	CF127034 UI-HF-ETO
18	763.6	60.0	1065	5	CF128484 UI-HF-ETO
19	752	59.1	818	4	BQ072411 AGENCOURT
20	751.2	59.1	1328	3	AK011575 Mus muscu
21	747.2	58.7	837	4	AK011575 Mus muscu
22	747	58.7	826	4	BI768711 603057274
C 23	745.4	58.6	877	7	CK770542 958729 MA
24	738.8	58.1	757	6	CF126927 UI-HF-ETO

25	734	57.7	744	6	CF128914	UI-HF-ETO
26	734	57.7	746	4	BI770998	BI770998 603055341
27	724.4	56.9	785	6	CF127896	CF127896 UI-HF-ETO
28	706.6	55.6	738	6	CF130905	CF130905 UI-HF-ETO
29	698.6	54.9	717	6	CF126467	CF126467 UI-HF-ETO
30	697.2	54.8	1167	4	BM811192	BM811192 AGENCOURT
31	696.8	54.8	718	7	CM409630	CM409630 170004245
32	683.8	53.8	852	7	CF994025	CF994025 AGENCOURT
C 33	681	53.5	696	5	BU677833	BU677833 UI-CF-ECO
C 34	673	52.9	677	5	BM975273	BM975273 UI-CF-EC1
35	670.4	52.7	779	4	BI837765	BI837765 603086450
C 36	668	52.5	815	2	AM025547	AM025547 xnu97f10.x
C 37	664.4	52.2	771	2	AM242648	AM242648 xnu10404.x
C 38	662.6	52.1	784	7	CK770983	CK770983 959214 MA
39	652	51.3	851	5	EX337613	EX337613
C 40	648.4	51.0	653	5	BQ016290	BQ016290 UI-H-DT1-
41	640	50.3	651	4	BI765033	BI765033 603051177
C 42	639	50.2	655	5	BQ019293	BQ019293 UI-H-DT1-
43	629.2	49.5	694	4	BI838855	BI838855 603082458
44	621.8	48.9	827	5	EX928295	EX928295 EX928295
45	621	48.8	924	4	BI763441	BI763441 603047447

#### ALIGNMENTS

RESULT 1  
CR616308  
LOCUS full-length cDNA clone CSODI070VK16 of Placentia Cot 25-normalized  
DEFINITION of Homo sapiens (human).  
ACCESSION CR616308  
VERSION CR616308.1 GI:50497115  
KEYWORDS HTC; CNSUT\_cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Li W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue  
Genoscope.  
2 (bases 1 to 1252)  
Direct Submission  
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
Location/Qualifiers  
1..1252  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI070VK16"  
/tissue\_type="Placentia Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 98.1%; Score 1248; DB 3; Length 1252;  
Best Local Similarity 100.0%; Pred. No. 1.2e-287;  
Matches 1248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCTACGCCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGTCGAGAGTC 60  
Db 5 TCTACGCCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGTCGAGAGTC 64

```
QY 61 CTGGGGTTCTTCTCACAACCTCTCGAAGGGGAAGGGTTGTGAGACCCACGAGACC 120
Db 65 CTGGGGTTCTTCTCACAACCTCTCGAAGGGGAAGGGTTGTGAGACCCACGAGACC 124
QY 121 CAACTCCAGTCCCGAGGAGGTGGCTGCGCACACTCGGAGGCGCTCTGGTTTCAGG 180
Db 125 CAACTCCAGTCCCGAGGAGGTGGCTGCGCACACTCGGAGGCGCTCTGGTTTCAGG 184
QY 181 GTCTCTCTCTCTCTCTCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 185 GTCTCTCTCTCTCTCTCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
QY 241 CTCTCTCCCTGTCCTACTCCCCCAACATGTCACCGGCTCCCTCAGCGATGTGAGGAC 300
Db 245 CTCTCTCCCTGTCCTACTCCCCCAACATGTCACCGGCTCCCTCAGCGATGTGAGGAC 304
QY 301 CTTCAAGAGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTCGAACAAGGAATTT 360
Db 305 CTTCAAGAGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTCGAACAAGGAATTT 364
QY 361 GTGACTTCCAAACAGAGACCGAGGAGACTCCAACTGCGAGAATGGTCTCCCGAGAG 420
Db 365 GTGACTTCCAAACAGAGACCGAGGAGACTCCAACTGCGAGAATGGTCTCCCGAGAG 424
QY 421 GCGCGGGCGGCTGGGCAAGAGAGAGAGGCGCCCAACGAAGAGCGCCCTGAGCGGG 480
Db 425 GCGCGGGCGGCTGGGCAAGAGAGAGAGGCGCCCAACGAAGAGCGCCCTGAGCGGG 484
QY 481 GTCAGCCAGAGGGGAAGAGTTCAGCGCAACGCGGCCCAACGCGGAGAGCGGCGCGC 540
Db 485 GTCAGCCAGAGGGGAAGAGTTCAGCGCAACGCGGCCCAACGCGGAGAGCGGCGCGC 544
QY 541 ATCGGAGTGTGAGCAAGGCTTCTCCAGACTCAAGACCAACCTCGCTGGGTGCCCGCC 600
Db 545 ATCGGAGTGTGAGCAAGGCTTCTCCAGACTCAAGACCAACCTCGCTGGGTGCCCGCC 604
QY 601 GACACCAAGTCTTCAAGCTGGACAGCTCAGGCTGGGCTCCAGTCTACATGCCCACTTG 660
Db 605 GACACCAAGTCTTCAAGCTGGACAGCTCAGGCTGGGCTCCAGTCTACATGCCCACTTG 664
QY 661 AGGAGATCTCTGCTTAACGACAATACGAGAGCGGTACATTCACCGGTCAACCTGAGG 720
Db 665 AGGAGATCTCTGCTTAACGACAATACGAGAGCGGTACATTCACCGGTCAACCTGAGG 724
QY 721 TGGCCCTTTATGTTGGCCGGGAACCGGAGAGTGACCTGAAAGAGTGTGACCGCGAGC 780
Db 725 TGGCCCTTTATGTTGGCCGGGAACCGGAGAGTGACCTGAAAGAGTGTGACCGCGAGC 784
QY 781 CGTTATGTGGAAACCAACCGCGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGGCTC 840
Db 785 CGTTATGTGGAAACCAACCGCGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGGCTC 844
QY 841 CCGGGGAGCGGCGCCCGGGAAGGAGCCCTGCGCTCAGTCTCTCTGCTCTGCTTCTTC 900
Db 845 CCGGGGAGCGGCGCCCGGGAAGGAGCCCTGCGCTCAGTCTCTCTGCTCTGCTTCTTC 904
QY 901 CCGCTCGCAATGCTCTCTCTGTCCTCCACCGCGGAGAACATTTTACAAACGAGGAG 960
Db 905 CCGCTCGCAATGCTCTCTCTGTCCTCCACCGCGGAGAACATTTTACAAACGAGGAG 964
QY 961 ATTGCTTCCAAACGAGAGATCAATGTGACTTACAAAGATTCCTCATCTATTAACTT 1020
Db 965 ATTGCTTCCAAACGAGAGATCAATGTGACTTACAAAGATTCCTCATCTATTAACTT 1024
QY 1021 TATTAACTTCTACGCTGAATGACTCTGCAAGCCTTGTGCTGCTCAAGTGAATATTAATT 1080
Db 1025 TATTAACTTCTACGCTGAATGACTCTGCAAGCCTTGTGCTGCTCAAGTGAATATTAATT 1084
QY 1081 ATAAATATATAATAGAGGCTTCAATGTACTTTTGTACAAATATGTGTAATAT 1140
Db 1085 ATAAATATATAATAGAGGCTTCAATGTACTTTTGTACAAATATGTGTAATAT 1144

1141 GTAGATCATAGGATAGCTGACTTTTGACAGTCACTTTATAAAGTAATTCACTTAAAGATA 1200
1145 GTAGATCATAGGATAGCTGACTTTTGACAGTCACTTTATAAAGTAATTCACTTAAAGATA 1204
1201 TATATTTTTTTTCAACAAGTTTGTCTACTTTTGAAATAAATCTTTCT 1248
1205 TATATTTTTTTTCAACAAGTTTGTCTACTTTTGAAATAAATCTTTCT 1252

RESULT 2
CR606057 1250 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1039J05 of Placentia Cot 25-normalized
of Homo sapiens (human)
ACCESSION CR606057 GI:50486864
VERSION CR606057.1 GI:50486864
KEYWORDS HTC; CNSLT_CDNA..
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1250)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1250)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source 1..1250
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1039J05"
/tissue_type="Placentia Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 97.5%; Score 1240; DB 3; Length 1250;
Best Local Similarity 100.0%; Pred. No. 1e-285;
Matches 1240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTACGCCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGTCGAGAGATC 60
Db 11 TCTACGCCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGTCGAGAGATC 70
QY 61 CTCGGGGTTCCTTCTCAAACTCTGCGAAGGGGAAGGGTTGTGAGACCCACGAGACC 120
Db 71 CTCGGGGTTCCTTCTCAAACTCTGCGAAGGGGAAGGGTTGTGAGACCCACGAGACC 130
QY 121 CAATCCAGTCTCCAGCAGGAGGTGGCTGCGCCACACTCGGAGGCGCTCTGTTTCAGG 180
Db 131 CAATCCAGTCTCCAGCAGGAGGTGGCTGCGCCACACTCGGAGGCGCTCTGTTTCAGG 190
QY 181 GTCTCTCTCTCTCTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 191 GTCTCTCTCTCTCTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 250
QY 241 CTCTCTCCCTGCTCACTCCCGCAACATGTCCAACCGGCTCCCTCAGCGATGTGAGGAC 300
Db 251 CTCTCTCCCTGCTCACTCCCGCAACATGTCCAACCGGCTCCCTCAGCGATGTGAGGAC 310
```









source	1.1044	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="CS01039YJ05"	
		/tissue_type="PLACENTA COT 25-NORMALIZED"	
		/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"	
		/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
ORIGIN			
Query Match	75.8%;	Score 963.8;	DB 1; Length 1044;
Best Local Similarity	97.9%;	Pred. No. 1.3e-219;	
Matches 1000;	Conservative	7; Mismatches	9; Indels 5; Gaps 3;
QY	1	TCACGGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTGCTCAGAGTC	60
Db	11	TCACGGCCACGACTCTGGAGTGGGAAACAGAGACCGGTTCTCTGCTCAGAGTC	70
QY	61	CTCGGGTTCTTCTCACAACTCTGCGAAGGGGAAAGGTTGTGAGACCAACAGACCC	120
Db	71	CTCGGGTTCTTCTCACAACTCTGCGAAGGGGAAAGGTTGTGAGACCAACAGACCC	130
QY	121	CAACTCAGCTCCACAGAGAGTGCTGGCCACACTCGGAGGCGCTTGGTTTCAGG	180
Db	131	CAACTCAGCTCCACAGAGAGTGCTGGCCACACTCGGAGGCGCTTGGTTTCAGG	190
QY	181	GTCTCTGCTCTCTCTCACCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCT	240
Db	191	GTCTCTGCTCTCTCTCACCTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCT	250
QY	241	CT	300
Db	251	CT	310
QY	301	CTTCAAGAGTGAGATGTTGAATGTGACGGTGTGAATGTGAATGTGAATGTGA	360
Db	311	CTTCAAGAGTGAGATGTTGAATGTGACGGTGTGAATGTGAATGTGAATGTGA	370
QY	361	GTGACTTCCAAAGAGACCCAGAGAGAGCTTCAACTGGGAGATGGGTTCTCCCAAG	420
Db	371	GTGACTTCCAAAGAGACCCAGAGAGAGCTTCAACTGGGAGATGGGTTCTCCCAAG	430
QY	421	GGCGGGGGCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	480
Db	431	GGCGGGGGCTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	490
QY	481	GTGAGCCAG	540
Db	491	GTGAGCCAG	550
QY	541	ATGCGAGTGTGAG	600
Db	551	ATGCGAGTGTGAG	610
QY	601	GACACCAAGCTCTCCAAAGTGGACAGCTCAGGCTGGGCTTCAGGTTACATCGCC	660
Db	611	GACACCAAGCTCTCCAAAGTGGACAGCTCAGGCTGGGCTTCAGGTTACATCGCC	670
QY	661	AGCAGATCTGCTTAACGACAAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	720
Db	671	AGCAGATCTGCTTAACGACAAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	730
QY	721	TGCGCCCTTTATGTGGCGCGGAAACCGGAGAGTACCTGAAAGAGTGTGACCGGAGC	780
Db	731	TGCGCCCTTTATGTGGCGCGGAAACCGGAGAGTACCTGAAAGAGTGTGACCGGAGC	790
QY	781	CGTTATGTGGAAACCAACCGCGTCTCTGACCTTGGAGGTGGAGTCTGGGAAAGCGCGT	840
Db	791	CGTTATGTGGAAACCAACCGCGTCTCTGACCTTGGAGGTGGAGTCTGGGAAAGCGCGT	850
QY	841	CCGGGGGAGCGGGCCCGGGAGCGACCCCTCCCTCAGTGTCTCTCTCTCTCTCTCT	900
Db	851	CCGGGGGAGCGGGCCCGGGAGCGACCCCTCCCTCAGTGTCTCTCTCTCTCTCTCT	910
QY	901	CCCTCGCAATGCT	960
Db	911	CCCTCGCAATGCT	969
QY	961	ATTGCTTTCCAAACAGAGAGATCAATTGTACTTACAAAGATCCCATCTATTAACTT	1020
Db	970	ATTGCTTTCCAAACAGAGAGATCAATTGTACTTACAAAGATCCCATCTATTAACTT	1025
QY	1021	T 1021	
Db	1026	W 1026	
RESULT 8			
AL552084/c			
LOCUS			
DEFINITION	AL552084 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA	1169 bp	linear
ACCESSION	AL552084		
VERSION	AL552084.3		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (Bases 1 to 1169)		
TITLE	Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	On Feb 15, 2001 this sequence version replaced gi:31273900.		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3812.f		
	For more information about this cluster, see		
	http://www.genoscope.cns.fr/cdna?s=CS0D1059CD09NP1&e=3812.f.		
FEATURES			
source	Location/Qualifiers		
	1..1169		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0D1059YH17"		
	/tissue_type="PLACENTA COT 25-NORMALIZED"		
	/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"		
	/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Query Match	75.5%;	Score 959.8;	DB 1; Length 1169;
Best Local Similarity	97.3%;	Pred. No. 1.2e-218;	
Matches 988;	Conservative	4; Mismatches	21; Indels 2; Gaps 2;
QY	207	TCCTCGGTTTCT	266
Db	1019	TCCTCGGTTTCT	961
QY	267	CATGTCCACCGGTCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGATGTTGGAATG	326
Db	960	CATGTCCACCGGTCCTCAGCGATGTGGAGACCTTCAAGAGTGGAGATGTTGGAATG	901
QY	327	TGACGGGTTGAAATGATTCGAACAGGAATTTGTGACTTCCACAGAGACCCGAGGA	385



```

Db 900 TGACGGTTGAAATGATTCGAACAAGAAATTTGTAATCTCAACGAGCAACCGAGA 841
Qy 387 GAGCTCAACTGCGAGAAATGGTCTCTCCAGAAAGGCGCGGCGCTGGGCAAGAGAG 446
Db 840 GAGCTCAACTGCGAGAAATGGTCTCTCCAGAAAGGCGCGGCGCTGGGCAAGAGAG 781
Qy 447 GAAGGCGCCACCAAGAAGAGCCCTGAGCGGGGTACACAGGAGGGGAACAGAGTCCA 506
Db 780 GAAGGCGCCACCAAGAAGAGCCCTGAGCGGGGTACACAGGAGGGGAACAGAGTCCA 721
Qy 507 GCGCAACGCGGCCAAACGCGGAGAGCGGCGCGCATGCGAGTGTGAGCAAGGCTTCTC 566
Db 720 GCGCAACGCGGCCAAACGCGGAGAGCGGCGCGCATGCGAGTGTGAGCAAGGCTTCTC 661
Qy 567 CAGACTCAAGACCAACCTCGCTGGTGGTCCCGCCGACACCAAGCTCTCCAACTGGACAC 626
Db 660 CAGACTCAAGACCAACCTCGCTGGTGGTCCCGCCGACACCAAGCTCTCCAACTGGACAC 602
Qy 627 GCTCAGGCTGGCTCCAGCTACATCGCCCACTGAGGAGATCCTGGCTAACGACAAATA 686
Db 601 GCTCAGGCTGGCTCCAGCTACATCGCCCACTGAGGAGATCCTGGCTAACGACAAATA 542
Qy 687 CGAAGCGGTATATTACCGGTCACCTGAGCTGCGGCTTTATGGTGGCGGGAAC 746
Db 541 CGAAGCGGTATATTACCGGTCACCTGAGCTGCGGCTTTATGGTGGCGGGAAC 482
Qy 747 CGAGAGTGACCTCAAGAAGATGTGACCGGAGCGCTTATGTGGAACCAACCGCTCTG 806
Db 481 CGAGAGTGACCTCAAGAAGATGTGACCGGAGCGCTTATGTGGAACCAACCGCTCTG 422
Qy 807 ACCTTGAGGTGCGAGTCTGGGAAAGCGGCTCCCGGGGAGCGGCGCCCGGAAGGC 866
Db 421 ACCTTGAGGTGCGAGTCTGGGAAAGCGGCTCCCGGGGAGCGGCGCCCGGAAGGC 362
Qy 867 GACCCCTGCGCTCAGTGCTCTGTCTGTCTGCTTCCCGCTCGCAATGCTCTCTCTGTC 926
Db 361 GACCCCTGCGCTCAGTGCTCTGTCTGTCTGCTTCCCGCTCGCAATGCTCTCTCTGTC 302
Qy 927 CCACCGCGGAGAACATTTACACGAGGAGATGCTTTCCAAACACAGAGGAGATCA 986
Db 301 CCACCGCGGAGAACATTTACACGAGGAGATGCTTTCCAAACACAGAGGAGATCA 242
Qy 987 ATTGTACTTACAGATTCCTATCTATTAACTTTTAACTTCTACCGTGGAATGACTCT 1046
Db 241 ATTGGACTTACAAAGATTCCTATCTATTAACTTTTAACTTCTACCGTGGAATGACTCT 182
Qy 1047 GCAAGCCTTGTGCTGCTCAAGTCAATATGTAATTAATAATAATAATAATAAGAGCCT 1106
Db 181 GCAAGCCTTGTGCTGCTCAAGTCAATATGTAATTAATAATAATAATAATAAGAGCCT 122
Qy 1107 ATCAATGTATCTTTGTACAAATGTTGTAATGTAATGTAATGTAATGTAATGTAATGTA 1166
Db 121 ATCAATGTATCTTTGTACAAATGTTGTAATGTAATGTAATGTAATGTAATGTAATGTA 62
Qy 1167 CAGTCAATTTATAAAGTAAATTCACCTTAAGATATATATTTTTTCAAAACAAGTT 1221
Db 61 CAGTCAATTTATAAAGTAAATTCACCTTAAGATATATATTTTTTCAAAACAAGTT 7

```

## RESULT 9

```

AL552380
LOCUS AL552380 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1070YK16 5-PRIME, mRNA sequence.
ACCESSION AL552380
VERSION AL552380.3 GI:45857168
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 965)

```

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:31274195.

Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0D1070BF08QF1&c=3812.f.

## FEATURES

source 1..965  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1070YK16"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /notes="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 73.8%; Score 939; DB 1; Length 965;  
 Best Local Similarity 99.8%; Pred. No. 1.le-213;  
 Matches 961; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
 Qy 1 TCTACGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTGCTGAGAGTTC 60  
 Db 5 TCTACGCCACGACTCTGGAGTGGGAAACAGAGAGCGGTTCTCTCTGCTGAGAGTTC 64  
 Qy 61 CTGGGGTTCCTTCTCACTCTCGAGAGGGAAGGTTGTGAGACCCCAACACGACCC 120  
 Db 65 CTGGGGTTCCTTCTCACTCTCGAGAGGGAAGGTTGTGAGACCCCAACACGACCC 124  
 Qy 121 CAATCCAGCTCCAGCAGAGGTGGCTGCGGCACACTCGGAGGCTCTTGGTTTCAG 180  
 Db 125 CAATCCAGCTCCAGCAGAGGTGGCTGCGGCACACTCGGAGGCTCTTGGTTTCAG 184  
 Qy 181 GTCTCTGTCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 Db 185 GTCTCTGTCT 244  
 Qy 241 CT 300  
 Db 245 CT 304  
 Qy 301 CTTCAAGAGGTGAGATGTTGGAATGTGACGGGTTGAAAATGGAATGGAATGGAATG 360  
 Db 305 CTTCAAGAGGTGAGATGTTGGAATGTGACGGGTTGAAAATGGAATGGAATGGAATG 364  
 Qy 361 GTGACTTCCAAACGAGAGCAGCGGAGAGCTCCAACTCGGAGAAATGGGTCTCCCCAGAAG 420  
 Db 365 GTGACTTCCAAACGAGAGCAGCGGAGAGCTCCAACTCGGAGAAATGGGTCTCCCCAGAAG 424  
 Qy 421 GGCCGCGCGCGCTGGGCAAGAGAGAGAGCGGCCCAACCAAGAGAGCCCTCTGAGGGGG 480  
 Db 425 GGCCGCGCGCGCTGGGCAAGAGAGAGAGAGCGGCCCAACCAAGAGAGCCCTCTGAGGGGG 484  
 Qy 481 GTCCAGCAGAGGAGGAGCGAGTCCAGCGCAACCGCGCCCAACCGCGAGAGCGGCCCGC 540  
 Db 485 GTCCAGCAGAGGAGGAGCGAGTCCAGCGCAACCGCGCCCAACCGCGAGAGCGGCCCGC 544  
 Qy 541 ATCGAGTGTGTGAGCAAGGCGCTTCTCCAGACTCAAGACCAACCTGCGCTGGGTGCCCCCCC 600  
 Db 545 ATCGAGTGTGTGAGCAAGGCGCTTCTCCAGACTCAAGACCAACCTGCGCTGGGTGCCCCCCC 604



QY 601 GACACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGGCCACTTG 660  
 Db 605 GACACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGGCCACTTG 664  
 QY 661 AGGAGATCTGCTTACGACCAATACGAAACGGGTACATTACCGCGTCAACCTGACG 720  
 Db 665 AGGAGATCTGCTTACGACCAATACGAAACGGGTACATTACCGCGTCAACCTGACG 724  
 QY 721 TGGCCCTTTATGTTGGCGGGAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGAGC 780  
 Db 725 TGGCCCTTTATGTTGGCGGGAACCCGAGAGTCACTGAAAGAGTGGTGAACCGGAGC 783  
 QY 781 CGCTTATGTTGAACCAACCGCTCTGACCTTGAAGTGGAGTGGGAAAGCGCGCTC 840  
 Db 784 CGCTTATGTTGAACCAACCGCTCTGACCTTGAAGTGGAGTGGGAAAGCGCGCTC 842  
 QY 841 CCGGGGGAGCGGGCCCGGGAAGGAGCCCTGCCCTCAGTCTCTCTGCTCTCTCTCTC 900  
 Db 843 CCGGGGGAGCGGGCCCGGGAAGGAGCCCTGCCCTCAGTCTCTCTGCTCTCTCTCTC 902  
 QY 901 CCCTCGCAATGCT 960  
 Db 903 CCCTCGCAATGCT 962  
 QY 961 ATT 963  
 Db 963 ATT 965

RESULT 10  
 BX337699  
 LOCUS 1056 bp mRNA linear EST 08-APR-2004  
 DEFINITION BX337699 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CSODI048YE15 5'-PRIME, mRNA sequence.  
 ACCESSION BX337699  
 VERSION BX337699  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1056)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On May 2, 2003 this sequence version replaced gi:30345636.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by life technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 3812.f  
 For more information about this cluster, see  
 http://www.genoscope.cns.fr/cdna?s=CSODI048AC08QP1&c=3812.f.  
 Location/Qualifiers

FEATURES  
 source  
 1..1056  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI048YE15"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 ORIGIN

Query Match 73.2%; Score 931; DB 5; Length 1056;  
 Best Local Similarity 96.9%; Pred. No. 9.2e-212;  
 Matches 953; Conservative 14; Mismatches 14; Indels 2; Gaps 2;  
 QY 1 TCTAGGGCCACGACTCTGGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAGAAATC 60  
 Db 11 TCTAGGGCCACGACTCTGGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGCAGAAATC 70  
 QY 61 CTGCGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGAGCC 120  
 Db 71 CTGCGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGTTGTGAGACCCCAACAGAGCC 130  
 QY 121 CAACTCCAGCTCCCAAGAGAGAGTGGTGGCCCACTCTCGGAGAGCCCTCTTGGTTTCAAG 180  
 Db 131 CAACTCCAGCTCCCAAGAGAGAGTGGTGGCCCACTCTCGGAGAGCCCTCTTGGTTTCAAG 190  
 QY 181 GTCTCTCTGTCT 240  
 Db 191 GTCTCTCTGTCT 250  
 QY 241 CT 300  
 Db 251 CT 310  
 QY 301 CTTCAAGAGGTGGAGATGTTGGAATGTGACGGGTGGAATGGAATGGAATGGAATGGA 360  
 Db 311 CTTCAAGAGGTGGAGATGTTGGAATGTGACGGGTGGAATGGAATGGAATGGAATGGA 370  
 QY 361 GTGACTTCCAAACGAGAGCACCAGAGAGAGTCCAACTCCGAGAAATGGGTCTCCCCAGAG 420  
 Db 371 GTGACTTCCAAACGAGAGCACCAGAGAGAGTCCAACTCCGAGAAATGGGTCTCCCCAGAG 430  
 QY 421 GCGCGCGCGCGCTGGGCAAGAGAGAGAGAGCGGCCCAACAGAGAGAGAGAGAGAGAGAG 480  
 Db 431 GCGCGCGCGCGCTGGGCAAGAGAGAGAGAGAGCGGCCCAACAGAGAGAGAGAGAGAGAGAG 490  
 QY 481 GTGAGCCAGAGAGGGAAGCAGGTCCAGCGCAACCGCGCAACCGCGCAACCGCGCAACCG 540  
 Db 491 GTGAGCCAGAGAGGGAAGCAGGTCCAGCGCAACCGCGCAACCGCGCAACCGCGCAACCG 550  
 QY 541 ATGCGAGTGTGAGAAAGGCGCTTCTCAGACTCAAGACCAACCTGCGGTGGGTGCGGCCCG 600  
 Db 551 ATGCGAGTGTGAGAAAGGCGCTTCTCAGACTCAAGACCAACCTGCGGTGGGTGCGGCCCG 610  
 QY 601 GACACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGCCCACTTG 660  
 Db 611 GACACCAAGCTCTCCAAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGCCCACTTG 670  
 QY 661 AGGAGATCTCTGGCTAACGACAAATACGAGAACGGGTACATTACCGCGTCAACCTGACG 720  
 Db 671 AGGAGATCTCTGGCTAACGACAAATACGAGAACGGGTACATTACCGCGTCAACCTGACG 730  
 QY 721 TGGCCCTTTATGTTGGCGGGAACCCGAGAGTGAACCTGAAAGAGTGGTGAACCGGAGC 780  
 Db 731 TGGCCCTTTATGTTGGCGGGAACCCGAGAGTGAACCTGAAAGAGTGGTGAACCGGAGC 790  
 QY 781 CGCTTATGTTGAACCAACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840  
 Db 791 CGCTTATGTTGAACCAACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 850  
 QY 841 CCGGGGGAGCGGGCCCGGGAAGGAGCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCTCT 900  
 Db 851 CCGGGGGAGCGGGCCCGGGAAGGAGCCCTGCCCTCAGTCTCTCTCTCTCTCTCTCTCTCT 909  
 QY 901 CCCCTCGCAATGCT 960  
 Db 910 CCCCTCGCAATGCT 968  
 QY 961 ATTGCTTTCCAAACGAGAGGAGA 983  
 Db 969 WTCGTTTCCAAACGAGAGGAGTA 991





end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3812.f

For more information about this cluster, see <http://www.genoscope.cns.fr/cdna?cs=CS0D1067BH09QP1&c=3812.f>.

FEATURES  
Location/Qualifiers  
1..996  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1067Y018"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN		Query Match	61.9%	Score 788	DB 5	Length 996
		Best Local Similarity	97.3%	Pred. No. 1.6e-177	Indels	1
		Matches 806	Conservative	5	Mismatches 16	Gaps 1
QY	16	CTGGAGTGGGAAACAGAGAGCGGTTCTCTGCTGCAGAGTCTCTCGGGTTCCTTCT	75			
DB	1	CTGGAGTGGGAAACAGAGAGCGGTTCTCTGCTGCAGAGTCTCTCGGGTTCCTTCT	60			
QY	76	CACAACTCTCGAAGGGGAAAGGGTTGTGAGACCCAAACAGAGCCCAACTCCAGCTCCCA	135			
DB	61	CACAACTCTCGAAGGGGAAAGGGTTGTGAGACCCAAACAGAGCCCAACTCCAGCTCCCA	120			
QY	136	GCAGAGTGGCTCGGCCACACTCGGAGGCGCTCTTGGTTTCAGGGTCTCTCTGTCTCTC	195			
DB	121	GCAGAGGCGGCGGCCACACTCGGAGGCGCTCTTGGTTTCAGGGTCTCTCTGTCTCTC	180			
QY	196	TCTCACCTCTTCTCGCTTCTCTGTCTCTGTCT				

DB	61	CAGCAGAGGTGGTGGCGCCACACTGGGAGCGCTCTTGGTTTCAGGGTCTCTCTGTCTC	120
QY	194	TCTCTCACCTCTTCT	253
DB	121	TCTCTCACCTCTTCT	180
QY	254	CACTCCCCCAACATGTCCACCGGCTCCCTCAGCGATGTGAGGACCTTCAAGAGGTGG	313
DB	181	CACTCCCCCAACATGTCCACCGGCTCCCTCAGCGATGTGAGGACCTTCAAGAGGTGG	240
QY	314	AGATGTTGAATGTGACGGGTTGAAATGGATTCGAACAAAGAAATTTGTGACTTCAACG	373
DB	241	AGATGTTGAATGTGACGGGTTGAAATGGATTCGAACAAAGAAATTTGTGACTTCAACG	300
QY	374	AGAGCACCAGGAGAGCTCAACTGCGAGAATGGGTCTCCCGAAGGCGCGCGGCGGCGGCGG	433
DB	301	AGAGCACCAGGAGAGCTCAACTGCGAGAATGGGTCTCCCGAAGGCGCGCGGCGGCGGCGG	360
QY	434	TGGCGAAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	493
DB	361	TGGCGAAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	420
QY	494	GGAAAGAGGTCCAGCGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	553
DB	421	GGAAAGAGGTCCAGCGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	480
QY	554	GCAAGGCTTCTCAGACTCAAGACCACTGCGCTGCGCTCAGCTACATCGCCCACTTGGAGGAGATCTCTGG	613
DB	481	GCAAGGCTTCTCAGACTCAAGACCACTGCGCTGCGCTCAGCTACATCGCCCACTTGGAGGAGATCTCTGG	540
QY	614	CCAAGCTGGACAGCTCAGGCTGGCTGCGCTCAGCTACATCGCCCACTTGGAGGAGATCTCTGG	673
DB	541	CCAAGCTGGACAGCTCAGGCTGGCTGCGCTCAGCTACATCGCCCACTTGGAGGAGATCTCTGG	600
QY	674	CTAACGCAAAATACGAGAGCGGTATACATCCCGGCTCAACCTGAGCTGGCGGCTTTATGG	733
DB	601	CTAACGCAAAATACGAGAGCGGTATACATCCCGGCTCAACCTGAGCTGGCGGCTTTATGG	660
QY	734	TGGCGGGAACCGGAGGTGACTCAAGAGAGTGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGG	793
DB	661	TGGCGGGAACCGGAGGTGACTCAAGAGAGTGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGG	720
QY	794	CCACCGGCTCTGACCTTGGAGTGGCGGTCTCGGAAAGCGCGCTCCCGGCGGAGCGG	853
DB	721	CCACCGGCTCTGACCTTGGAGTGGCGGTCTCGGAAAGCGCGCTCCCGGCGGAGCGG	780
QY	854	GCCCCGGGAGGCGA 868	
DB	781	GCCCCGGGAGGCGA 795	

RESULT 14  
BX360244  
LOCUS  
DEFINITION BX360244 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1067Y018 5-PRIME, mRNA sequence.  
ACCESSION BX360244  
VERSION BX360244.2 GI:46553101  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 996)  
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30382426.  
Contact: Genoscope  
Genoscope - Centre National de Sequenage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

Qy	736	GC	CGGAAACCCGAGAGTGACCTGAAAGAGTGTGTGACCGGAGCGCGTATTGTGGAACC	795
Db	721	GC	CGGAAACCCGAGAGTGACCTGAAAGAGTGTGTGACCGGAGCGCGTATTGTGGAAC	779
Qy	796	AC	CGCGTCTCCTGACCTTGAGAGTGGGAGTCTCGGAAAGCGCGCTCCCG	843
Db	780	AC	CGCGTCTCCTGACCTTGAGAGTGGGAGTCTCGGAGCGAGGTTCTCCG	827
RESULT 15				
CFI27519				
LOCUS				
DEFINITION			800 bp mRNA linear EST 05-AUG-2003	
			UI-HF-ETO-awh-f-14-0-UI.r1 NIH_MGC_214 Homo sapiens CDNA clone	
			IMAGE:30554533 5', mRNA sequence.	
ACCESSION			CFI27519	
VERSION			CFI27519.1	
KEYWORDS			GI:33205841	
SOURCE			EST.	
ORGANISM			Homo sapiens (human)	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE			1 (Bases 1 to 800)	
			Bonaldo,M.F., Lennon,G. and Soares,M.B.	
			Normalization and subtraction: two approaches to facilitate gene	
			discovery	
JOURNAL			Genome Res. 6 (9), 791-806 (1996)	
MEDLINE			97044477	
PUBMED			8889548	
COMMENT			Contact: Soares, MB	
			Coordinated Laboratory for Computational Genomics	
			University of Iowa	
			375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA	
			Tel: 319 335 8250	
			Fax: 319 335 9565	
			Email: bento-soares@uiowa.edu	
			Tissue Procurement: Mary Hendrix	
			CDNA Library preparation: Dr. M. Bento Soares, University of Iowa	
			CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa	
			DNA Sequencing by: Dr. M. Bento Soares, University of Iowa	
			Clone Distribution: Distribution information can be found at	
			http://genome.uiowa.edu/distribution/humanfil.html	
			The following repetitive elements were found in this CDNA	
			sequence: 172-253, >(GA)n\$imple_repeat (matched complement)	
			Seq primer: pyx-5.	
FEATURES			Location/Qualifiers	
source			1..800	
			/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="taxon:9606"	
			/clone="IMAGE:30554533"	
			/tissue_type="Chondrosarcoma Lung Metastasis cell lines"	
			/lab_host="DH10B (T1 phage resistant)"	
			/clone_lib="NIH_MGC_214"	
			/note="Organ: Lung; Vector: pyx-Asc; Site 1: EcoR I;	
			Site 2: Not I; The library was constructed according	
			Bonaldo, Lennon and Soares, Genome Research, 6:791-806,	
			1996. Denatured RNA was size fractionated on a 1% agarose	
			gel. First strand cDNA synthesis was primed with oligo-dT	
			primer containing a Not I site. Double strand cDNA was	
			size selected according to mRNA size fraction, ligated	
			with EcoR I adaptor, digested with Not I and then cloned	
			directionally into pyx-Asc vector. The library tag	
			sequence located between the Not I site and the polyA tail	
			is GATTAAGGCCA. Tissue was provided by Mary Hendrix."	
ORIGIN				
Query Match			61.9%; Score 787; DB 6; Length 800;	
Best Local Similarity			100.0%; Pred. No. 2.6e-177;	
Matches			787; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	11	CG	ACTCTCGGAGTGGGAAACAGAGACCGGTTCTCTCTGTCGACGAAGTCTCGGGGTT	70
Db	1	CG	ACTCTCGGAGTGGGAAACAGAGACCGGTTCTCTCTGTCGACGAAGTCTCGGGGTT	60

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 21:52:41 ; Search time 88 Seconds

(without alignments)  
1445.809 Million cell updates/sec

Title: US-09-701-674A-23

Perfect score: 917

Sequence: 1 MSGSLSDVEDQEVEMLEEC.....PESDLKEVTSRLCGTTAS 179

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh  
-Q/cgn2\_1/USPTO.spool\_h/US09701674/runat\_15122004.100441.27529/app\_query.fasta\_1.327  
-DB=Issued Patents NA -QFWT=fastcap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 @CGN 1.1.69 @runat\_15122004.100441.27529 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptcdat/1/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptcdat/1/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptcdat/1/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptcdat/1/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptcdat/1/ina/PTCUS\_COMB.seq:\*

6: /cgn2\_6/ptcdat/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	38.5	240	US-09-016-434-1028	Sequence 1028, Ap
2	184	17.9	1800	US-09-771-357-106	Sequence 106, App
3	154	16.8	1275	US-08-552-142A-3	Sequence 3, Appli
4	154	16.8	1275	US-08-910-973-3	Sequence 3, Appli
5	154	16.8	1275	US-09-499-227-3	Sequence 3, Appli
6	154	16.8	1275	PCT-US95-05741-3	Sequence 3, Appli
7	146.5	16.0	1560	US-08-552-142A-14	Sequence 14, Appl
8	146.5	16.0	1560	US-08-910-973-14	Sequence 14, Appl
9	146.5	16.0	1560	US-09-499-227-14	Sequence 14, Appl
10	146.5	16.0	2502	US-09-234-332-1	Sequence 1, Appli
11	146	15.9	2089	US-08-552-142A-1	Sequence 1, Appli
12	146	15.9	2089	US-08-910-973-1	Sequence 1, Appli

13	146	15.9	2089	4	US-09-499-227-1	Sequence 1, Appli
14	146	15.9	2089	5	PCT-US95-05741-1	Sequence 1, Appli
15	145.5	15.9	1462	1	US-08-552-142A-16	Sequence 16, Appl
16	145	15.8	1951	1	US-08-910-973-16	Sequence 16, Appl
17	145	15.8	1951	4	US-09-499-227-16	Sequence 16, Appl
18	143.5	15.6	1527	4	US-08-722-570-12	Sequence 12, Appl
19	143.5	15.6	1527	4	US-08-932-411A-12	Sequence 12, Appl
20	143	15.6	1352	1	US-08-552-142A-10	Sequence 10, Appl
21	143	15.6	1385	4	US-08-932-411A-17	Sequence 17, Appl
22	143	15.6	1535	1	US-08-910-973-10	Sequence 10, Appl
23	143	15.6	1535	4	US-09-499-227-10	Sequence 10, Appl
24	143	15.6	1550	3	US-09-234-332-3	Sequence 3, Appli
25	140.5	15.3	738	4	US-08-722-570-13	Sequence 13, Appl
26	140.5	15.3	738	4	US-08-932-411A-13	Sequence 13, Appl
27	140.5	15.3	1333	1	US-08-910-973-21	Sequence 21, Appl
28	140.5	15.3	1333	4	US-09-499-227-21	Sequence 21, Appl
29	140.5	15.3	1676	3	US-09-234-332-2	Sequence 2, Appli
30	140	15.3	5340	4	US-09-535-145-1	Sequence 1, Appli
31	138.5	15.1	524	1	US-08-552-142A-8	Sequence 8, Appli
32	138.5	15.1	524	1	US-08-910-973-8	Sequence 8, Appli
33	138.5	15.1	524	4	US-09-499-227-8	Sequence 8, Appli
34	138.5	15.1	524	5	PCT-US95-05741-8	Sequence 8, Appli
35	137	14.9	804	4	US-08-932-411A-19	Sequence 19, Appl
36	137	14.9	1861	4	US-09-535-145-3	Sequence 3, Appli
37	131	14.3	310	1	US-08-552-142A-12	Sequence 12, Appl
38	129.5	14.1	485	5	PCT-US95-05741-10	Sequence 10, Appl
39	126	13.7	1268	1	US-08-910-973-12	Sequence 12, Appl
40	126	13.7	1268	4	US-09-499-227-12	Sequence 12, Appl
41	125	13.6	1312	4	US-08-722-570-14	Sequence 14, Appl
42	125	13.6	1312	4	US-08-932-411A-15	Sequence 15, Appl
43	120	13.1	1238	4	US-09-774-528-350	Sequence 350, App
44	119	13.0	1277	4	US-08-722-570-15	Sequence 15, Appl
45	119	13.0	1277	4	US-08-932-411A-16	Sequence 16, Appl

## ALIGNMENTS

### RESULT 1

US-09-016-434-1028  
; Sequence 1028, Application US/09016434  
; Patent No. 6500398

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/016,434

; FILING DATE: HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Zeller, Karen J.

; REGISTRATION NUMBER: 37,071

; REFERENCE/DOCKET NUMBER: PA-0002 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555





```

; ORIGINAL SOURCE:
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..1083
; US-08-552-142A-3
;
; Alignment Scores:
; Pred. No.: 3,64e-09 Length: 1275
; Score: 154.00 Matches: 67
; Percent Similarity: 46.46% Conservative: 25
; Best Local Similarity: 33.84% Mismatches: 64
; Query Match: 16.79% Indels: 43
; DB: 1 Gaps: 8
;
; US-09-701-674A-23 (1-179) x US-08-552-142A-3 (1-1275)
;
; QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24
;
; Db 103 AGTTCTCAGGATGAACACGATCTGGAGAA----- 132
;
; QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerAsnCys 44
;
; Db 133 -----AAGGAGGAGAGATGATGAAGAAGACGATGAAGACTCACTGAATCAT 180
;
; QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg--- 59
;
; Db 181 CACATGGAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGCGATGA 239
;
; QY 60 -----AAGGAGGAGAGATGATGAAGAAGACGATGAAGACTCACTGAATCAT 180
;
; Db 240 AGATGATGATGAGGATGACGACCAACCCAGAACCCGAGGAGGAGGAGGAGGAGGAG 239
;
; QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89
;
; Db 300 GACGAAAGCCCGGTGGAGGATTTAAAGTGAGACGATGAAGGCAACGCCAGGAGAG 359
;
; QY 89 gAlaArgMetArgValLeuSerLysAlaPheSerArgLysThrThrLeuProTirpVa 109
;
; Db 360 GAATCGCATGCGGACTCAACGATCCCTGACAGCTGCGCAAGTGTCGCCCTGCTA 419
;
; QY 109 lProProAspThrLysLeuSerLysLeuAspThrLysLeuArgLeuAlaSerSerTyrIleAl 129
;
; Db 420 CTCGAAACACAAAGTTGCTTAAGATTGAAACTCTGCGCTGGCTAAGAACTACATCTG 479
;
; QY 129 ahLysLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsn-----GlyTyrIleHisPr 147
;
; Db 480 GCCTCTTCTGAGATTTAAGTTCGCGCAAAAGCCGACCTGCTGCTCTTGTGTACAAC 539
;
; QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159
;
; Db 540 TCTCTGAAAGGTTTCTGCGAGCCACCAACCAATCTAGTAGCGGGTGTCGACGTGAA 599
;
; QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174
;
; Db 600 CCCAGAACTTTCCTTCTGAGAGATCCAGACATCCAGTCCGACATGC 649
;
; RESULT 4
; US-08-910-973-3
; Sequence 3, Application US/06910973
; Patent No. 5795723
;
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
;
; Alignment Scores:
; Pred. No.: 3,64e-09 Length: 1275
; Score: 154.00 Matches: 67
; Percent Similarity: 46.46% Conservative: 25
; Best Local Similarity: 33.84% Mismatches: 64
; Query Match: 16.79% Indels: 43
; DB: 1 Gaps: 8
;
; US-08-910-973-3 (1-179) x US-08-910-973-3 (1-1275)
;
; QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24
;
; Db 103 AGTTCTCAGGATGAACACGATCTGGAGAA----- 132
;
; QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerAsnCys 44
;
; Db 133 -----AAGGAGGAGAGATGATGAAGAAGACGATGAAGACTCACTGAATCAT 180
;
; QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg--- 59
;
; Db 181 CACATGGAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGCGATGA 239
;
; QY 60 -----AAGGAGGAGAGATGATGAAGAAGACGATGAAGACTCACTGAATCAT 180
;
; Db 240 AGATGATGATGAGGATGACGACCAACCCAGAACCCGAGGAGGAGGAGGAGGAGGAG 239
;
; QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89
;
; Db 300 GACGAAAGCCCGGTGGAGGATTTAAAGTGAGACGATGAAGGCAACGCCAGGAGAG 359
;
; QY 89 gAlaArgMetArgValLeuSerLysAlaPheSerArgLysThrThrLeuProTirpVa 109
;
; Db 360 GAATCGCATGCGGACTCAACGATCCCTGACAGCTGCGCAAGTGTCGCCCTGCTA 419
;
; QY 109 lProProAspThrLysLeuSerLysLeuAspThrLysLeuArgLeuAlaSerSerTyrIleAl 129
;
; Db 420 CTCGAAACACAAAGTTGCTTAAGATTGAAACTCTGCGCTGGCTAAGAACTACATCTG 479
;
; QY 129 ahLysLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsn-----GlyTyrIleHisPr 147
;
; Db 480 GCCTCTTCTGAGATTTAAGTTCGCGCAAAAGCCGACCTGCTGCTCTTGTGTACAAC 539
;
; QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159
;
; Db 540 TCTCTGAAAGGTTTCTGCGAGCCACCAACCAATCTAGTAGCGGGTGTCGACGTGAA 599
;
; QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174
;
; Db 600 CCCAGAACTTTCCTTCTGAGAGATCCAGACATCCAGTCCGACATGC 649
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/910,973
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA: WO PCT/US95/05741
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FPCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1275 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..1083
; US-08-910-973-3
;
; Alignment Scores:
; Pred. No.: 3,64e-09 Length: 1275
; Score: 154.00 Matches: 67
; Percent Similarity: 46.46% Conservative: 25
; Best Local Similarity: 33.84% Mismatches: 64
; Query Match: 16.79% Indels: 43
; DB: 1 Gaps: 8
;
; US-09-701-674A-23 (1-179) x US-08-910-973-3 (1-1275)
;
; QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24
;
; Db 103 AGTTCTCAGGATGAACACGATCTGGAGAA----- 132
;
; QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerAsnCys 44
;
; Db 133 -----AAGGAGGAGAGATGATGAAGAAGACGATGAAGACTCACTGAATCAT 180
;
; QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg--- 59
;
; Db 181 CACATGGAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGCGATGA 239
;
; QY 60 -----AAGGAGGAGAGATGATGAAGAAGACGATGAAGACTCACTGAATCAT 180
;
; Db 240 AGATGATGATGAGGATGACGACCAACCCAGAACCCGAGGAGGAGGAGGAGGAGGAG 239
;
; QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89
;
; Db 300 GACGAAAGCCCGGTGGAGGATTTAAAGTGAGACGATGAAGGCAACGCCAGGAGAG 359
;
; QY 89 gAlaArgMetArgValLeuSerLysAlaPheSerArgLysThrThrLeuProTirpVa 109
;
; Db 360 GAATCGCATGCGGACTCAACGATCCCTGACAGCTGCGCAAGTGTCGCCCTGCTA 419

```

QY 109 lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAl 129  
 Db 420 CTCACAAACACAAAGTTGCTAAGATTGAACCTTCGCCCTGGCTAAGAACTACATCG 479  
 QY 129 aHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsn-----GlyTyrIleHisPr 147  
 Db 480 GGCTCTTTCTGAGATTTAAGGTCGGCAAAAGCCAGACCTGGTGTCTCTTTGTACAAAC 539  
 QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159  
 Db 540 TCTCTGCAAAAGTTTCTGCAGCCCAACCAATCTAGTAGCGGGGTGCTCCAGCTGAA 599  
 QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174  
 Db 600 CCCAGAACTTCTCTCTGAGCAGAGTCAGGACATCCAGTCGCACATGC 649

## RESULT 5

US-09-499-227-3  
 ; Sequence 3, Application US/09499227  
 ; Patent No. 644463  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tapscott, Stephen J.  
 ; APPLICANT: Olson, James M.  
 ; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
 ; STREET: 1420 Fifth Avenue, Suite 2800  
 ; CITY: Seattle  
 ; STATE: WA  
 ; COUNTRY: USA  
 ; ZIP: 98101-2347  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/499,227  
 ; FILING DATE: 05-August-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/239,238  
 ; FILING DATE: 06-May-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: WO PCT/US95/05741  
 ; FILING DATE: 08-May-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US96/17532  
 ; FILING DATE: 30-October-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/910,973  
 ; FILING DATE: 07-August-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sheiness, Diana K.  
 ; REGISTRATION NUMBER: 35,356  
 ; REFERENCE/DOCKET NUMBER: FHC-1-12742  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
 ; TELEFAX: 206-225-0779  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1275 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Xenopus laevis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 25..1083  
 ; US-09-499-227-3

Alignment Scores:  
 Pred. No.: 3 64e-09 Length: 1275  
 Score: 154.00 Matches: 67  
 Percent Similarity: 46.46% Conservative: 25  
 Best Local Similarity: 33.84% Mismatches: 64  
 Query Match: 16.79% Indels: 43  
 DB: 4 Gaps: 8

US-09-701-674A-23 (1-179) x US-09-499-227-3 (1-1275)

QY 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuGluCysAspGlyLeuLys 24  
 Db 103 AGTTCTCAGGATGAAAACGATCTCGAGAA----- 132  
 QY 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerSerAsnCys 44  
 Db 133 -----AAGGAGGAGAGATTGATGAAAGAGACGATGAAGACTCAGTGAATCAT 180  
 QY 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg--- 59  
 Db 181 CACAATCGAGAGGA-GAACGAGGAGAGGATGAAGGGGATGAGGAGGAGGAGGAGCATGA 239  
 QY 60 -----ArgLysAlaProThr-LysLysSerProLe 69  
 Db 240 AGATGATGATGAGATGACACGACAGAAACCCAAAGGCGAGGCCGAAAGAAAAAAT 299  
 QY 69 userGlyValSerGlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluAr 89  
 Db 300 GACGAAAGCCCGGTGGAGCGATTAAAGTGAGACGCATGAAGGCAACGCCACGAGGAG 359  
 QY 89 GAlaArgMetArgValLeuSerLysLysAlaPheSerArgLeuLysThrLeuProTrpVa 109  
 Db 360 GAATCGCATGCGGACTCAACGATCCCTGGACACTCTCGCAAAAGTTGTGCCCTGCTA 419  
 QY 109 lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAl 129  
 Db 420 CTCACAAACACAAAGTTGCTAAGATTGAACCTTCGCCCTGGCTAAGAACTACATCTG 479  
 QY 129 aHisLeuArgGlnIleLeuAlaAsnAspLysTyrGluAsn-----GlyTyrIleHisPr 147  
 Db 480 GGCTCTTTCTGAGATTTAAGGTCGGCAAAAGCCAGACCTGGTGTCTCTTTGTACAAAC 539  
 QY 147 oVal-----AsnLeuThrTrpProPhe-----MetValAlaGly-----Ly 159  
 Db 540 TCTCTGCAAAAGTTTCTGCAGCCCAACCAATCTAGTAGCGGGGTGCTCCAGCTGAA 599  
 QY 159 s-ProGlu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174  
 Db 600 CCCAGAACTTCTCTCTGAGCAGAGTCAGGACATCCAGTCGCACATGC 649

## RESULT 6

PCT-US95-05741-3  
 ; Sequence 3, Application PC/TUS9505741  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weintraub, Harold  
 ; APPLICANT: Lee, Jacqueline B.  
 ; APPLICANT: Tapscott, Stephen J.  
 ; APPLICANT: Hollenberg, Stanley M.  
 ; TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Gene  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Christensen O'Connor Johnson Kindness  
 ; STREET: 1420 Fifth Avenue, Suite 2800  
 ; CITY: Seattle  
 ; STATE: WA  
 ; COUNTRY: USA  
 ; ZIP: 98101-2347  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25

600 CCCGAACTTTCCTCTCTAGCAGAGTCCAGACATCCAGCATGCCACATGC 68

Db

RESULT 7  
US-08-552-142A-14  
Sequence 14, Application US/08552142A  
Patent No. 5695995  
GENERAL INFORMATION:  
APPLICANT: Weintraub, Harold M.  
APPLICANT: Lee, Jacqueline E.  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Hollenberg, Stanley M.  
TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes  
TITLE OF INVENTION: and Proteins  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/552,142A  
FILING DATE: 02-Nov-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FHCR-1-8933  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-225-0709  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1560 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: HC2A  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 57..1126  
US-08-552-142A-14

Alignment Scores:  
Pred. No.: 4,16e-08 Length: 1560  
Score: 146.50 Matches: 41  
Percent Similarity: 50.37% Conservativeness: 27  
Best Local Similarity: 30.37% Mismatches: 66  
Query Match: 15.98% Indels: 1  
DB: 1 Gaps: 1

US-09-701-674A-23 (1-179) x US-08-552-142A-14 (1-1560)

Qy 6 LeuSerAspValGluAspLeuGlnValGluMetLeuCysAspGlyLeu 1560  
Db 135 CTCAGTTCTCAGGACGACGAGCAAGAAGGAGGAGCGACCTTCGA 1111

Db

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/05741  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FHCR-1-8504  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-225-0709  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Xenopus laevis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 25..1083  
PCT-US95-05741-3

Alignment Scores:  
Pred. No.: 3,64e-09 Length: 1275  
Score: 154.00 Matches: 67  
Percent Similarity: 46.46% Conservativeness: 25  
Best Local Similarity: 33.84% Mismatches: 64  
Query Match: 16.79% Indels: 43  
DB: 5 Gaps: 8

US-09-701-674A-23 (1-179) x PCT-US95-05741-3 (1-1275)

Qy 5 SerLeuSerAspValGluAspLeuGlnGluValGluMetLeuCysAspGlyLeuLys 24  
Db 103 AGTCTCAGATGAACAACATCTGGAAA-----132

Qy 25 MetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGluSerSerAsnCys 44  
Db 133 -----AAGGAGGAGGTGTGATGAAGAAGAGCATGAAGACTCACTCAATCAT 180

Qy 45 GluAsnGlySerProGlnLysGlyArgGly-----GlyLeuGlyLysArg--- 59  
Db 181 CACATGGAGGA-GAACGAGAGAGAGATGAAGGGATGAGGAGGAGGAGGAGTGTA 239

Qy 60 -----ArgLysAlaProThr-LysLysSerProLe 69  
Db 240 AGATGATGATGAGATGACGACCAGAAACCCAAAGCGGAGCGMAAAAAAATAAT 299

Qy 69 uSerGlyValSerGlnGluGlyGlnValGlnArgAsnAlaLaenAlaArgGluAr 89  
Db 300 GAGCAAAGCCGGTGGACCGATTAAAGTAGAGCGCATGAAGGCAACGCCAGGAGAG 359

Qy 89 GalArgMetArgValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTrpVa 109  
Db 360 GAATCGCATGCAAGGACTCAACGATGCCCTGGACAGTCTGGCAAGTGTGCTCTGCTA 419

Qy 109 lProProAspThrLysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrlleAl 129  
Db 420 CTCCAAACACAAAAGTTCTTAAGATTGAATACTTCGCGCTGGCTAAGAACTACATCTG 479

Qy 129 aHisLeuArgGlnIleLeuAlaAsnAspLysTyrgLuAsn-----GlyTyrlleHisPr 147  
Db 480 GGCTCTTCTGAGATTTTAAAGTTCGGCAAAAGCCAGACCTGGTGTCTTTGACAAAC 539

Qy 147 oVal-----AsnLeuThrTrpProPhe-----MetValalagly-----Ly 159  
Db 540 TCTTCGAAAGGTTTCTTCGACCCACCAACCAATCTAGTAGCGGGGTCTCTGACGTGAA 599

Qy 159 s-Proglu---SerAspLeuLysGluValValThrAlaSerArgLeuCys 174





;; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
;; STREET: 1420 Fifth Avenue, Suite 2800  
;; CITY: Seattle

Alignment Scores:		
Pred. No.:	7,498-08	Length:
Score:	146.00	Matches:
Percent Similarity:	49.23%	Conservative:
Best Local Similarity:	34.63%	Mismatches:
Query Match:	15.93%	Indels:
DB:	1	Gaps:
		3







```

Qy 143 -----GlyTyrIleHisProVal-AsnLeuThrTrpProPheMetV 156
Db 797 CGTGCAGACTCTGTGCAAGGGGCTGTACAGCCGCCACGAAATCTGGTGGCGGCTGCCT 856
Qy 156 alAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyT 176
Db 857 GCAGT-----TAACTCTCGTAACTTCTCTCAGGAGCGGGCC 895
Qy 176 hrThrAlaSer 179
Db 896 GGACGGGGGCC 906

```

Search completed: December 17, 2004, 00:05:29  
 Job time : 94 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 16, 2004, 20:16:52 ; Search time 455 seconds

(without alignments)  
2065.157 Million cell updates/sec

Title: US-09-701-674a-23

Perfect score: 917

Sequence: 1 MGTGSLSDVDELQVEMLLEC.....PESDLKVVVTASRLCGTTAS 179

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4134886 segs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool.h/US09701674/runat\_15122004\_100440\_27494/app\_query.fasta\_1.327  
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701674 -CGN\_1\_1\_470 -runat\_15122004\_100440\_27494 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04: \*  
1: Geneseq1980s: \*  
2: Geneseq1990s: \*  
3: Geneseq2000s: \*  
4: Geneseq2001as: \*  
5: Geneseq2001bs: \*  
6: Geneseq2002as: \*  
7: Geneseq2002bs: \*  
8: Geneseq2003as: \*  
9: Geneseq2003bs: \*  
10: Geneseq2003cs: \*  
11: Geneseq2003ds: \*  
12: Geneseq2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	917	100.0	1272	3	Aaz57861
2	764.5	83.4	632	6	ABQ55427
3	514	56.1	2177	10	AD225693
4	514	56.1	2196	3	AAJ72428
5	514	56.1	2196	10	ADJ56314
6	507	55.3	2382	6	AAS94828

7	495.5	54.0	1718	3	Aaz50464
8	493.5	53.8	1716	3	Aaz50465
9	487	53.1	482	9	ACH21857
10	353	38.5	240	10	ACA56430
11	353	38.5	240	12	ADJ56226
12	318	34.7	697	6	ABL65519
13	318	34.7	697	6	ABL65935
14	318	34.7	697	6	ABK64784
15	228	24.9	493	9	ACH23969
16	211.5	23.1	1065	4	ABL08723
17	208	22.7	5101	4	ABL08722
18	185	20.2	963	8	ABZ18589
19	183.5	20.0	1276	10	ADC30883
20	183.5	20.0	1395	12	ADQ25304
21	183.5	20.0	1406	8	ACC62520
22	183.5	20.0	1406	8	ACC62491
23	180.5	19.7	1140	6	ABQ88217
24	180.5	19.7	1160	10	ADC32623
25	174.5	19.0	981	10	ADJ51583
26	169	18.4	474	5	AAS76438
27	168.5	18.4	1004	6	ABK71581
28	165.5	18.0	651	4	AAH48230
29	165.5	18.0	651	4	AAH49596
30	165.5	18.0	651	4	AAH44361
31	165.5	18.0	2466	8	ABX3251
32	164	17.9	609	12	ACH89545
33	164	17.9	797	11	ADP65591
34	164	17.9	1396	11	ADP65052
35	164	17.9	1664	12	ADQ22344
36	164	17.9	1800	6	ABT05509
37	164	17.9	1800	11	ADM83716
38	164	17.9	2870	12	ADQ17684
39	162	17.7	531	4	ABL16071
40	162	17.7	1457	6	ABL63712
41	162	17.7	1457	11	ADP65592
42	162	17.7	1678	6	AAS94864
43	157.5	17.2	907	9	AAI62234
44	157.5	17.2	2537	6	ABSS1803
45	157.5	17.2	2550	10	ADD29657

## ALIGNMENTS

RESULT 1

Aaz57861

ID Aaz57861 standard; cDNA; 1272 BP.

XX AC Aaz57861;

XX AC Aaz57861;

DT 11-APR-2000 (first entry)

XX DE Protein regulating gene expression PRGE-23 cDNA clone 3340296.

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

XX DE Protein regulating gene expression; PRGE-23; human; cancer; inflammation;

(INCY-) INCYTE PHARM INC.  
 XX Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;  
 PI Guegler KJ, Gorgone GA, Baughn WR, Patterson C, Lu DM;  
 XX WPI; 2000-116543/10.  
 DR P-PSDB; AAY58630.  
 XX New human polypeptides that regulate gene expression, for treatment,  
 PT prevention and diagnosis of, e.g. cancer.  
 PT  
 XX Claim 9; Page 139; 150pp; English.  
 PS  
 XX The present sequence is that of Incyte clone 3340296 encoding new human  
 CC protein regulating gene expression PRGE-23 (see AAY58630). The cDNA was  
 CC initially isolated from spleen tissue cDNA library SPLNNOT10, and the  
 CC full-length sequence assembled from overlapping clones from a number of  
 CC libraries. PRGE-23 is expressed in reproductive, developmental and  
 CC urologic tissues associated with cancer, inflammation and foetal  
 CC diseases, disorders or conditions. It is characterised as an Myc-type HLH  
 CC protein. The invention provides PRGE polypeptides (see AAY58608-38) and  
 CC polynucleotides (see AAZ57839-69), expression vectors, host cells,  
 CC antibodies, agonists and antagonists. It also provides methods for  
 CC diagnosing, treating or preventing disorders associated with expression  
 CC of PRGE. Polynucleotides are also used as sources of probes and primers  
 CC for diagnosis and monitoring of disease, also for detecting related  
 CC sequences and in gene mapping  
 XX  
 SQ Sequence 1272 BP; 311 A; 358 C; 318 G; 285 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 4,19e-92 Length: 1272  
 Score: 917.00 Matches: 1272  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0  
 US-09-701-674A-23 (1-179) x AAZ57861 (1-1272)  
 QY 1 MetSerThrGlySerLeuSerAspValGluAppLeuGlnGluValMetLeuGluCys 20  
 DB 268 ATGTCACCGGCTCCCTCAGCATGTTGGAGGACCTTCAAGAGGTGGAGATGTGGAATGT 327  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 DB 328 GACGGGTTGAAATGGATTGCAACAGGAATTTGTGACTTCCACGAGAGCACCAGGAG 387  
 QY 41 SerSerAsnGlyGluAsnGlySerProGlnLysGlyArgGlyLeuGlyLysArgArg 60  
 DB 388 AGCTCCAACTGCGAGAAATGGTCTCCCAAGAGGGCGCGGCTGGCGCAAGAGGAGG 447  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyGlnValGln 80  
 DB 448 AAGGCGCCCAACCAAGAGAGCCCTTGAGCGGGGTGAGCCAGAGGGAGGAGGTCGAG 507  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSer 100  
 DB 508 CGCAACGCGCCCAACGCGGAGAGCGGGCCCGCATGCGAGTCTGAGCAAGGCTTCCTCC 567  
 QY 101 ArgLeuLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThr 120  
 DB 568 AGACTCAAGACCAACCTTGGCTGGTGGCGCCCGCCAGCACCAGCTCTCCAAAGTGGACACG 627  
 QY 121 LeuArgLeuAlaSerSerTrpIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 DB 628 CTGAGGCTGGGCTCAGCTACATCGCCCTCTGAGGCGAGATCTCTGCTTACACGAAATAC 687  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysPro 160  
 DB 688 GAGACGGGTGATATTACCGCGGTCAACCTGAGTGGCCCTTTATGTGGTGGCCGGGAACCC 747  
 QY 161 GluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179

Db 748 GAGAGTGACCTGAAGAAGTGTGACCGGAGCGCTTATGTGGACCAACCGCGTCC 804  
 RESULT 2  
 ABQ55427  
 ID ABQ55427 standard; cDNA; 632 BP.  
 XX AC ABQ55427;  
 XX DT 22-AUG-2002 (first entry)  
 XX DE Human ovarian antigen HNOJ185 cDNA, SEQ ID NO:1307.  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 OS Homo sapiens.  
 XX WO200200677-A1.  
 XX 03-JAN-2002.  
 XX 07-JUN-2001; 2001WO-US018569.  
 XX 07-JUN-2000; 2000US-0209467P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;  
 XX WPI; 2002-147878/19.  
 XX P-PSDB; ABP42350.  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.  
 PS Claim 1; SEQ ID NO 1307; 2922pp; English.  
 XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present

CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_sequences  
 XX  
 SQ Sequence 632 BP; 145 A; 176 C; 201 G; 93 T; 0 U; 17 Other;

Alignment Scores:  
 Pred. No.: 1.56e-75 Length: 632  
 Score: 764.50 Matches: 163  
 Percent Similarity: 91.62% Conservative: 1  
 Best Local Similarity: 91.06% Mismatches: 14  
 Query Match: 83.37% Indels: 3  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x AB055427 (1-632)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValMetLeuGluCys 20  
 Db 13 ATGTCCACCGGCTCCCTCAGCGATGTGGAGACCTTCAGAGGTGGAGATGTGGATGT 72  
 QY 21 AspGlyLeuLysMetAspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu 40  
 Db 73 GACGGNNTGAAATGGATTCCAAAGGAATTTGTGACTTCCACGAGAGCACCGAGGAG 132  
 QY 41 SerSerAsnCysGluAsnGlySerProGlnLysGlyArgGlyGlyLeuGlyLysAspArg 60  
 Db 133 AGTCTCAACTGCAGATGGTCTCCCAAGAGGCGCGGNGGCTTGGCGAAGAGGAG 192  
 QY 61 LysAlaProThrLysLysSerProLeuSerGlyValSerGlnGluGlyLysGlnValGln 80  
 Db 193 AAGCGGCCCAACCAAGAGAGCCCTCAGCGGTGTGAGNCAGAGGAGGAGAGGAGTNCAG 252  
 QY 81 ArgAsnAlaAlaAsnAlaArgGluArgAlaAactMetArgValLeuSerLysAlaPheSer 100  
 Db 253 CGCAACGCCGNCAGCGGAGAGGCGGNCCTGATGCGATGCTGTGACGAGGCTTCTCC 312  
 QY 101 ArgLeuLysThrThrLeuProThrValProAspThrLysLeuSerLysLeuAspThr 120  
 Db 313 AGACTCAAGACCACTGCTGCTGGTGGTCCCGCCGACCAAGCTCTCCAAGCTGGACAG 372  
 QY 121 LeuArgLeuAlaSerSerThrIleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyr 140  
 Db 373 CTGAGGTGGCGTCCAGCTACATCGCCCACTTGAGGACAGATCTGTGCTAACGACCAATAC 432  
 QY 141 GluAsnGlyTyrIleHisProValAsnLeuThrProPheMetValAlaGlyLysPro 160  
 Db 433 NAGAACGGGTACATTA--CCCGGCACTGACCTGNCCTTTATGTGTC--GGGAACCC 489  
 QY 161 GluSerAspLeuLysGluValVal-ThrAlaSerArgLeuCysGlyThrAla 178  
 Db 490 GAGAGTGANCTGAAGAAGTGTGTGACCGGAGCCGGTTATGTGGAANCAACCGCG 544

RESULT 3

ADP25693  
 ID ADE25693 standard; cDNA; 2177 BP.  
 XX  
 AC ADE25693;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human cDNA differentially expressed in foam cells #97.  
 XX  
 KW Human; ss; differential expression; foam cell; LPS; lipopolysaccharide;  
 XX cardiovascular disease; atherosclerosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003194721-A1.  
 XX  
 PD 16-OCT-2003.  
 XX  
 PF 18-SEP-2002; 2002US-00247671.

XX  
 PR 19-SEP-2001; 2001US-0323784P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Mikita T, Shiffman D, Porter JG, Kaser MR;  
 XX  
 DR WPI; 2003-875398/81.  
 XX  
 DR P-P8DB; ADE25769.  
 XX  
 PT Combination containing several polynucleotide that are differentially  
 PT expressed in foam cells and complements of the polynucleotides, useful  
 PT for diagnosing cardiovascular disease or atherosclerosis.  
 XX  
 PS Claim 1; SEQ ID NO 97; 37pp; English.  
 XX

The invention relates to a combination comprising several polynucleotides  
 having any one of 127 sequences (S1) such as the sequence of human  
 calmodulin gene, human mRNA for KIA0930 protein, leukotriene A4  
 hydrolase, human CGI-142 protein mRNA, human K+ channel beat 2 subunit  
 mRNA, etc., and their complements. The cDNAs are differentially expressed  
 in LPS (lipopolysaccharide)-treated foam cells. Also included are  
 obtaining an extended or full length gene from a library of nucleic acid  
 sequences, an expression vector containing the nucleic acids, a host cell  
 containing the vector, a purified polypeptide appearing as ADE25750 and  
 ADE25751, producing a protein by culturing the host cell, and a  
 composition comprising a purified antibody that specifically binds to the  
 proteins. The foam cell-expressed nucleic acids are useful for a high  
 throughput detection of differential expression of one or more  
 polynucleotides in a sample. The sample is from a subject with  
 atherosclerosis and comparison with a standard defines early, mid or late  
 stages of the disorder. The foam cell-expressed nucleic acids are useful  
 for high throughput screening of a library of molecules or compounds to  
 identify a ligand which binds a polynucleotide. The library is chosen  
 from DNA molecules, peptides, proteins and RNA molecules. The protein is  
 useful for a high throughput screening of library of molecules or  
 compounds to identify at least one ligand which specifically binds a  
 protein, for purifying a ligand from a sample for making an antibody. The  
 foam cell-expressed nucleic acids are useful for diagnosing  
 cardiovascular disorder. The foam cell-expressed nucleic acids are useful  
 as elements on a microarray which can be used for detecting related  
 polynucleotide in a sample, diagnosing cardiovascular disease,  
 atherosclerosis. The present sequence represents a cDNA whose expression  
 is upregulated in LPS treated foam cells.

SQ Sequence 2177 BP; 423 A; 691 C; 650 G; 413 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.99e-47 Length: 2177  
 Score: 514.00 Matches: 113  
 Percent Similarity: 63.21% Conservative: 21  
 Best Local Similarity: 53.30% Mismatches: 38  
 Query Match: 56.05% Indels: 40  
 DB: 10 Gaps: 5

US-09-701-674A-23 (1-179) x ADE25693 (1-2177)

QY 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValMetLeuGluCys 20  
 Db 458 ATGTCCACCGGCTCGGTGAGTGATCCGAGAG-----ATCGAGCTT 499  
 QY 21 AspGlyLeuLys-----Met 25  
 Db 500 CGGGGGCTGCAGCGGAGTACCCGCTCCCAAGAGCGGCCCTCCGCGCGGTA 559  
 QY 26 AspSerAsnLysGluPheValThrSerAsnGluSerThrGluGlu----- 40  
 Db 560 GAGCGCAGTACGCTCGCCCGAGTACACTCTCGCGCAGAGAGGAGGAGCCCGACGGC 619  
 QY 41 -----SerSerAsnCysGluAsnGlySer 48  
 Db 620 GAGGAGGAGCGCTCGCTCTGGGCACAGCGCGGAGCGGAGGCTCAAGAGGAGCGG 679





XX Composition useful for diagnosis of conditions, disorders or diseases  
 PT associated with atherosclerosis, comprises several polynucleotides that  
 PT are differentially expressed in foam cell development.  
 XX Claim 1; Page 135-136; 315pp; English.  
 XX The present invention relates to the isolation of human polynucleotide  
 CC sequences that are differentially expressed during foam cell  
 CC differentiation. The polynucleotide sequences of the invention or a  
 CC composition comprising these polynucleotides are useful as a high  
 CC throughput method for detecting altered expression of one or more  
 CC polynucleotides in a sample. The polynucleotides can be used in the  
 CC diagnosis of disorders associated with foam cell development such as  
 CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as  
 CC coronary artery disease. The polynucleotide sequences can also be used as  
 CC PCR primers and probes. The polynucleotides of the invention are also  
 CC useful in gene therapy. AAS94746-AAS95021 represent the human  
 CC polynucleotide sequences of the invention which are differentially  
 CC expressed during foam cell differentiation  
 XX SQ Sequence 2382 BP; 459 A; 732 C; 702 G; 489 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4,77e-46 Length: 2382  
 Score: 507.00 Matches: 115  
 Percent Similarity: 63.13% Conservative: 22  
 Best Local Similarity: 53.00% Mismatches: 31  
 Query Match: 55.29% Indels: 49  
 DB: 6 Gaps: 6

US-09-701-674a-23 (1-179) x AAS94828 (1-2382)

Qy 1 MetSerThrGlySerLeuSerAspValGluAspLeuGlnGluValMetLeuGluCys 20  
 Db ATGTCACGGGCTCGGTGAGTATCGGAGGAG-----ATGGAGCTT 502  
 Qy 21 AspGlyLeuLys-----Met 25  
 Db CGGGGGCTGCGGGGAGTACCGGTCCCGCCCTCCCAAGAGCGCGCCCTCCGGCGGCTGA 562  
 Qy 26 AspSerAsnLysGluPheValThrSerAsnGluSerThrGlu----- 39  
 Db GAGCGCAGCTACCGCTCGCCAGTGAACCTCGTCGCGAGAGGAGGAGGAGCCCGCAGCGC 622  
 Qy 40 GluSerSerAsnCysGluLysGlnGlySerProGlnLysGlyArgGly----- 54  
 Db GAGGAGGAGCGCTCGCTCTGGGCACACGCGCGCAGCGCGGAGAGGCTTCCAGAGAGCGG 682  
 Qy 55 -----GlyLeuGlyLysArgArgLys-----Ala 62  
 Db CCCCCGTGCTGGGGCGCGCGCGAGGTGTACCGCGCGCGGTGGTGGCAAGAGGCC 742  
 Qy 63 ProThrLysLysSerProLeuSerGlyValSerGln-GluGlyLysGlnValGlnArgAs 82  
 Db CCGCGCCCAAGGGCTCAGCGCGAGAGTGCAGACAGTCCGACGCGGAA 790  
 Qy 82 nAlaAlaAsnAlaArgGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLe 102  
 Db CGCGGCCCAAGCGCGGTGAGGTGCCCGATGCGGTGTGAGCAAGCCCTTCCAGGCT 850  
 Qy 102 uLysThrThrLeuProTrpValProProAspThrLysLeuSerLysLeuAspThrLeuAr 122  
 Db CAAGACCCAGCTCGCTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 910  
 Qy 122 gLeuAlaSerSerTyrlleAlaHisLeuArgGlnIleLeuAlaAsnAspLysTyGluAs 142  
 Db GCTGGCTTCAGTATACCTCTCACCTGCGGAGCTGTTCAGAGGAGCGGCTATGAGAA 970  
 Qy 142 nGlyTyrlleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluLe 162  
 Db CGGCTAGCTGACCCAGTGAACCTGACATGCCATTCTCGTGGTCTCGGGAAGACCGGACTC 1030

Qy 162 rAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla 178  
 Db TGACACCAAGAAAGTTTCCGACGCAACAGACTATGTGGAACCAACCGCT 1079  
 RESULT 7  
 AAZ50464  
 ID AAZ50464 standard; cDNA; 1718 BP.  
 AC AAZ50464;  
 DT 23-MAY-2000 (first entry)  
 XX Murine muscudin bHLH protein encoding cDNA.  
 DE Muscudin; basic helix-loop-helix; bHLH; transcription factor; myogenesis;  
 KW screening; myogenic disease; aberration; muscle development; cytotatic;  
 KW gene therapy; diagnostic agent; muscular dystrophy; myopathy;  
 KW neuromuscular; skeletomuscular; myogenic cancer; mouse; ss.  
 OS Mus sp.  
 XX FH Key Location/Qualifiers  
 FT CDS 219..824  
 FT /\*tag= a  
 FT /product= "Murine muscudin bHLH protein"  
 FT /note= "Lacks Alanine-Threonine myogenic recognition motif"  
 XX WO200006720-A1.  
 XX 10-FEB-2000.  
 XX 30-JUL-1999; 99WO-AU000623.  
 XX 30-JUL-1998; 98AU-00004955.  
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.  
 PI Robb L, Begley CG, Harvey RP;  
 DR WPI; 2000-195290/17.  
 DR P-PSDB; XAY44921.  
 XX New regulatory protein useful for gene therapy of myogenic cancer and for  
 PT detecting aberrations in muscle development in myogenic disease,  
 PT comprises protein associated with genes involved in myogenesis.  
 XX Claim 3; Fig 1A; 54pp; English.  
 CC The present cDNA sequence encodes murine muscudin bHLH (basic helix-loop-  
 CC helix) protein, which is a transcription factor associated with genes  
 CC involved in myogenesis. Muscudin gene can be used in genetic screening  
 CC for myogenic disease conditions, for detecting aberrations in muscle  
 CC development and in gene therapy. Anti-muscudin antibodies can be used as  
 CC therapeutic or diagnostic agents. Muscudin has cytotatic activity and  
 CC can be used to treat myogenic disease conditions like myopathies,  
 CC muscular dystrophies, neuromuscular and skeletomuscular disorders and  
 CC myogenic cancer  
 XX SQ Sequence 1718 BP; 356 A; 477 C; 520 G; 365 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 5.8e-45 Length: 1718  
 Score: 495.50 Matches: 112  
 Percent Similarity: 63.90% Conservative: 19  
 Best Local Similarity: 54.63% Mismatches: 43  
 Query Match: 54.03% Indels: 31  
 DB: 3 Gaps: 5

US-09-701-674A-23 (1-179) x AAZ50464 (1-1718)

Qy 1 MetSerThrGlySerLeuSerAspValGluAsp-----LeuGlnGluVal 15  
 |||||



```

Db 219 ATGTCCACCGGCTCGGTGAGTGACCCGGAAGACTCGAGATGAGGGGCTCGAGAGGCTC 278
Qy 16 GluMetLeuGluCysAspGly-----LeuLysMetAspSerAsnLysGluPhe 31
Db 279 TACCGGCGCCGCTCCAGAGCGCGCCCTGCTCCGATGAGCGGGTTACGGCTCG 338
Qy 32 ValThrSerAsnGluSerThrGluGluSerSerAsnCysGluAsnGlySerProGlnLys 51
Db 339 CCAGGCGACATTTCTTCTCGGAAGAG-----GAGGACGGTGAAGAGAGGCC 386
Qy 52 Gly-----ArgGly 54
Db 387 GGTCTCCCTGGAGCCCGGGAGATGCAAGAGAAAGCGGCTCCGTGGGGCTGACCTGGC 446
Qy 55 GlyLeuGlyLysArgArgLysAlaProThrLysLysSer---ProLeuSerGlyValSer 73
Db 447 GCGCAGGTGGCGCGCAGCGGTGCGGGGAAAGCCGCTCCCGCTTAAGGCTCGGCC 506
Qy 74 GlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArg 93
Db 507 GCAGAGTGAAGCAGTCGACGCGGAATGCGGCAACGCCCGGAAACGCGCCGGATGGCC 566
Qy 94 ValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTyrValProProAspThr 113
Db 567 GTGCTGAGCAAGACCTTCTCCAGACTGAAGACAGGCTGCCCTGGTGGCCGCCACACC 626
Qy 114 LysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGln 133
Db 627 AAGCTTTCCAACTGACACGCTGCGCTGGCTTCCAGCTACATCGCGCACCTGGCGCAG 686
Qy 134 IleLeuAlaAsnAspLysTyrGluAsnGlyTyrIleHisProValAsnLeuThrPro 153
Db 687 CTGCTCGAGAGACCGCTACGAGGACAGCTATGTGCACCCCTGTGAACCTGACGTGGCCA 746
Qy 154 PheMetValAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSerArgLeu 173
Db 747 TTCGTGCTCTGGAGCGCCAGACTCTGACACAAAGAGCTTCTTCGACCAACAGGCTT 806
Qy 174 CysGlyThrThrAla 178
Db 807 TGTGGAACCTTCGCT 821

RESULT 8
AAZ50465
ID AAZ50465 standard; cDNA; 1716 BP.
XX AC AAZ50465;
XX DT 23-MAY-2000 (first entry)
XX DE Human musculin bHLH protein encoding cDNA.
XX KW Musculin; basic helix-loop-helix; bHLH; transcription factor; myogenesis;
XX screening; myogenic disease; aberration; muscle development; cyostatic;
XX KW gene therapy; diagnostic agent; muscular dystrophy; myopathy;
XX KW neuromuscular; skeletomuscular; myogenic cancer; human; ss.
XX OS Homo sapiens.
XX FH Key
XX FT CDS
XX FT Location/Qualifiers
XX FT 2..607
XX FT /*tag= a
XX FT /product= "Human musculin bHLH protein"
XX FT /note= "Lacks Alanine-Threonine myogenic recognition
XX FT motif"
XX PN WO200006720-A1.
XX PD 10-FEB-2000.
XX PP 30-JUL-1999; 99NO-AU000623.
XX PR 30-JUL-1998; 98AU-00004955.

```

```

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
PA Robb L, Begley CG, Harvey RP;
PI WPI; 2000-195290/17.
DR P-ESDB; AAY44922.
XX New regulatory protein useful for gene therapy of myogenic cancer and for
XX detecting aberrations in muscle development in myogenic disease,
XX comprises protein associated with genes involved in myogenesis.
XX Example 3; Page 38-40; 54pp; English.
XX The present cDNA sequence encodes human musculin bHLH (basic helix-loop-
XX helix) protein, which is a transcription factor associated with genes
XX involved in myogenesis. Musculin gene can be used in genetic screening
XX for myogenic disease conditions, for detecting aberrations in muscle
XX development and in gene therapy. Anti-musculin antibodies can be used as
XX therapeutic or diagnostic agents. Musculin has cyostatic activity and
XX can be used to treat myogenic disease conditions like myopathies,
XX muscular dystrophies, neuromuscular and skeletomuscular disorders and
XX myogenic cancer
XX SQ Sequence 1716 BP; 336 A; 550 C; 518 G; 312 T; 0 U; 0 Other;

Alignment Scores:
Pred. NO.: 9.67e-45 Length: 1716
Score: 493.50 Matches: 102
Percent Similarity: 70.30% Conservative: 14
Best Local Similarity: 61.82% Mismatches: 30
Query Match: 53.82% Indels: 19
DB: 3 Gaps: 3

US-09-701-674A-23 (1-179) x AAZ50465 (1-1716)
Qy 33 ThrSerAsnGluSerThrGluGlu----- 40
Db 110 AGTGACAACTCGTCCGACAGGAGGAGGACCCCGAGGAGGAGGAGCGCTCGCTCTG 169
Qy 41 -----SerSerAsnCysGluAsnGlySerProGln---LysGlyArgGly 54
Db 170 GGCACAGCGCGAGCGCGGAAGGCTGCAAGAGGAAGCGCGCGCGGGCTGGGGCGGC 229
Qy 55 GlyLeuGlyLysArgArgLysAlaProThrLysLysSer---ProLeuSerGlyValSer 73
Db 230 GCGCAGGTGGTAGCGGGCGGTGGTGGCAAGAGCCCTCCCGCCAAAGGCTCAGCC 289
Qy 74 GlnGluGlyLysGlnValGlnArgAsnAlaAlaAsnAlaArgGluArgAlaArgMetArg 93
Db 290 GCAGAGTGCAAGCAGTCGACGCGGACGCGGCAACGCCCGTGAGCGTGCCTCGCGATCGC 349
Qy 94 ValLeuSerLysAlaPheSerArgLeuLysThrThrLeuProTyrValProProAspThr 113
Db 350 GTGTGAGCAAGACCTTCTCCAGCTCAGACGACGCTGCCCTGGTGGTGGTGGTGGTGGTGGT 409
Qy 114 LysLeuSerLysLeuAspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGln 133
Db 410 AAGCTCTCCAGCTGGACAGCTCCGCTGGCTTCCAGTTACATCGCTACCTCGCGCAG 469
Qy 134 IleLeuAlaAsnAspLysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpPro 153
Db 470 CTGTGTGAGGAGGACCGCTATGAGAACGGCTAGCTGACCCAGTGAACCTGACATGGCCA 529
Qy 154 PheMetValAlaGlyLysProGluSerAspLeuLysGluValValThrAlaSerArgLeu 173
Db 530 TTCGTGCTCTGGGAAGACCGGACTCTGACACCAAGAGTTTCCGCGAGCGACAGACTA 589
Qy 174 CysGlyThrThrAla 178
Db 590 TGTGGAACCGCT 604

RESULT 9

```

ACH21857  
 ID ACH21857 standard; cDNA; 482 BP.  
 XX AC ACH21857;  
 XX DT 13-OCT-2003 (first entry)  
 XX DE Human adult ovary cDNA #237.  
 XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 XX KW genome mapping; biodiversity; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN US2003073623-A1.  
 XX PD 17-APR-2003.  
 XX PF 30-JUL-2001; 2001US-00918995.  
 XX PR 30-JUL-2001; 2001US-00918995.  
 XX PA (DRMA/) DRMANAC R T.  
 XX PA (LABA/) LABAT I.  
 XX PA (STAC/) STACHE-CRAIN B.  
 XX PA (DICK/) DICKSON M C.  
 XX PA (JONE/) JONES L W.  
 XX PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
 XX DR WPI; 2003-615964/58.  
 XX PT New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.  
 XX PS Claim 1; SEQ ID NO 9069; 44pp; English.  
 XX CC The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH21857-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
 XX SQ Sequence 482 BP; 99 A; 156 C; 143 G; 83 T; 0 U; 1 Other;  
 Alignment Scores:  
 Pred. No.: 8,84e-45 Length: 482  
 Score: 487.00 Matches: 95  
 Percent Similarity: 98.96% Conservative: 0  
 Best Local Similarity: 98.96% Mismatches: 1  
 Query Match: 53.11% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-701-674A-23 (1-179) x ACH21857 (1-482)  
 QY 84 AlaAsnAlaGluArgAlaArgMetArgValLeuSerLysAlaPheSerArgLeuLys 103  
 Db 33 GCNCATAGCGGAGAGCGGCGCCGATCGAGTGCTGAGCAGGCGCTTCCAGACTCAAG 92

QY 104 ThrThrLeuProTrpValProAspThrLysLeuSerLysLeuAspThrLeuArgLeu 123  
 Db 93 ACCACCTGCCCCCTGGGTGCCCCCGACCAAGCTCTCCAAGCTGGACACGCTCAGGCTG 152  
 QY 124 AlaSerSerTrpIleAlaHisLeuArgGlnIleAlaAsnAspLysTrpGluAsnGly 143  
 Db 153 GCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACGCAAAATACGAGAACGGG 212  
 QY 144 TyrIleHisProValAsnLeuThrTrpProPheMetValAlaGlyLysProGluSerAsp 163  
 Db 213 TACATTCAACCGGTCAACCTGACGTGGCCCTTTATGGTGGCCGGAAACCCGAGATGAC 272  
 QY 164 LeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAlaSer 179  
 Db 273 CTGAAGAGAGTGGTGACCGGAGCGGCTTATGTGAACCAACCGCGTCC 320  
 RESULT 10  
 ID ACA56430 standard; cDNA; 240 BP.  
 XX AC ACA56430;  
 XX DT 06-JUN-2003 (first entry)  
 XX DE Chicken signalling pathway polynucleotide probe SEQ ID NO 1028.  
 XX KW Chicken; probe; ss; array element; Parkinson's disease;  
 KW signalling pathway population; cancer; adenocarcinoma; leukaemia;  
 KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.  
 XX OS Gallus gallus.  
 XX PN US6500938-B1.  
 XX PD 31-DEC-2002.  
 XX PF 30-JAN-1998; 98US-00016434.  
 XX PR 30-JAN-1998; 98US-00016434.  
 XX PA (INCY-) INCYTE GENOMICS INC.  
 XX PI Au-Young J, Seilhamer JJ;  
 XX DR WPI; 2003-352189/33.  
 XX PT Combination of polynucleotide probes, useful as array elements in a  
 PT microarray for monitoring the expression of a number of target  
 PT polynucleotides.  
 XX PS Claim 1; SEQ ID NO 1028; 65pp; English.  
 XX CC The invention relates to a combination which, comprises a number of  
 CC polynucleotide probes comprising a sequence selected from one of the 1490  
 CC sequences mentioned in the specification. The combination is useful as an  
 CC array element in a microarray for monitoring the expression of a number  
 CC of target polynucleotides. The microarray is particularly useful in the  
 CC diagnosis and treatment of cancer and immunopathology and neuropathology.  
 CC The microarray is useful in diagnostics and treatment regimens, drug  
 CC discovery and development, toxicological and carcinogenicity studies,  
 CC forensics and pharmacogenomics. The microarray is also useful for  
 CC monitoring progression of diseases and for developing sophisticated  
 CC profiles for the effects of currently available therapeutic drugs. The  
 CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs  
 CC and genomic fragments and in research and diagnostic applications. The  
 CC array can detect changes in expression in a large number of genes coding  
 CC for different signalling pathway populations which can be used to diagnose  
 CC various diseases including cancer e.g. adenocarcinoma and leukaemia,  
 CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease  
 CC and Parkinson's disease. The present sequence represents a polynucleotide  
 CC probe of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format directly from USPTO at



XX 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0233133P.  
PR 18-SEP-2000; 2000US-0233617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234924P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 28-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237295P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.  
PA (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX Claim 1; SEQ ID NO 3856; 44pp; English.  
XX The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 847 sequences (given in ABL6164  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophagus, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms  
CC cell carcinoma.

CC tumour  
XX Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
SQ Alignment Scores:  
Pred. No.: 9.92e-26 Length: 697  
Score: 318.00 Matches: 61  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 34.68% Indels: 0  
DB: 6 Gaps: 0  
US-09-701-674A-23 (1-179) x ABL65519 (1-697)  
QY 119 AspThrLeuArgLeuAlaSerSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnApp 138  
Db 3 GACACGCTCAGGTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACGAC 62  
QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProPheMetValAlaGly 158  
Db 63 AATACGAGAGACGGGTACATTCACCGGTCACTGACGTGGCCCTTATGTGGACACCGCG 122  
QY 159 LysProGluSerAspLeuLysGluValThrAlaSerArgLeuCysGlyThrThrAla 178  
Db 123 AAACCCGAGAGTGACCTGAAAGAGTGGTGACCGCGCGCTTATGTGGACACCGCG 182  
QY 179 Ser 179  
Db 183 TCC 185  
RESULT 13  
ABL65935  
ID ABL65935 standard; DNA; 697 BP.  
XX ABL65935;  
AC ABL65935;  
XX 15-MAY-2002 (first entry)  
XX Lung cancer related gene sequence SEQ ID NO:4272.  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilms's tumour; adenocarcinoma;  
KW gene; ds.  
XX Homo sapiens.  
XX WO200194629-A2.  
XX 13-DEC-2001.  
XX 30-MAY-2001; 2001WO-US010838.  
XX 05-JUN-2000; 2000US-0209473P.  
XX 05-JUN-2000; 2000US-0209531P.  
XX 18-SEP-2000; 2000US-0233133P.  
XX 18-SEP-2000; 2000US-0233617P.  
XX 20-SEP-2000; 2000US-0234009P.  
XX 20-SEP-2000; 2000US-0234034P.  
XX 20-SEP-2000; 2000US-0234052P.  
XX 22-SEP-2000; 2000US-0234509P.  
XX 22-SEP-2000; 2000US-0234567P.  
XX 25-SEP-2000; 2000US-0234923P.  
XX 25-SEP-2000; 2000US-0234924P.  
XX 25-SEP-2000; 2000US-0235077P.  
XX 25-SEP-2000; 2000US-0235082P.  
XX 25-SEP-2000; 2000US-0235134P.  
XX 25-SEP-2000; 2000US-0235280P.  
XX 26-SEP-2000; 2000US-0235637P.  
XX 26-SEP-2000; 2000US-0235638P.  
XX 27-SEP-2000; 2000US-0235711P.  
XX 27-SEP-2000; 2000US-0235720P.  
XX 27-SEP-2000; 2000US-0235840P.

PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236942P.  
 PR 29-SEP-2000; 2000US-0236991P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVAL-) AVALON PHARM.  
 XX  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 DR WPI; 2002-188264/24.  
 XX  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 XX Claim 1; SEQ ID NO 4272; 44pp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 847 sequences (given in ABU61664  
 CC to ABU70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms'  
 CC tumour  
 XX  
 SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;

Alignment Scores:  
 Pred. No.: 9.92e-26 Length: 697  
 Score: 318.00 Matches: 61  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 34.68% Indels: 0  
 DB: 6 Gaps: 0

US-09-701-674A-23 (1-179) x ABL65935 (1-697)

QY 119 AspThrLeuArgLeuAlaSerTyrIleAlaHisLeuArgGlnIleLeuAlaAsnAsp 138  
 DB 3 GACACGCTAGGCTGGCGTCCACCTACATCGCCCACTTGAGGAGATCTGCTAACGAC 62  
 QY 139 LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpPheMetValAlaGly 158  
 DB 63 AAATACGAGAACGGGTACATTACCCCGGTCAACCTGACGTGGCCCTTTATGTGGCGCGG 122

QY 159 LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrAla 178  
 DB 123 AAACCCGAGAGTACCTGAAGAAGTGGTGACCGGAGCGCTTATGTGAACACCGCG 182  
 QY 179 Ser 179  
 DB 183 TCC 185  
 RESULT 14  
 ABK64784  
 ID ABK64784 standard; DNA; 697 BP.  
 XX  
 AC ABK64784;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Human benign prostatic hyperplasia gene #679.  
 XX  
 KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200212440-A2.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 07-AUG-2001; 2001WO-US024708.  
 XX  
 PR 07-AUG-2000; 2000US-0223232P.  
 PR 05-JUN-2001; 2001US-00873319.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 PA (NLSB) JAPAN TOBACCO INC.  
 XX  
 PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
 XX  
 WPI; 2002-257476/30.  
 XX  
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
 PT detecting expression levels of one or more genes in prostate cells from  
 PT patient that are differentially regulated compared to normal prostate  
 XX cells.  
 PS Disclosure; Page 367; 44pp; English.  
 XX  
 CC The invention relates to a method of diagnosing (I) the onset or  
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for  
 CC or identifying an agent that modulates the onset or progression of BPH.  
 CC The method is based on changes in gene expression in BPH tissue isolated  
 CC from patients exhibiting different clinical states of prostate  
 CC hyperplasia as compared to normal prostate tissue. (I) comprises  
 CC detecting the expression levels of one or more genes in prostate cells  
 CC from the subject that are differentially regulated compared to normal  
 CC prostate cells. (II) comprises preparing a first gene expression profile  
 CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention  
 XX  
 SQ Sequence 697 BP; 190 A; 164 C; 156 G; 180 T; 0 U; 7 Other;  
 Alignment Scores:  
 Pred. No.: 9.92e-26 Length: 697

Score:	318.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	34.68%	Indels:	0
DB:	6	Gaps:	0

US-09-701-674A-23 (1-179) X ABK64784 (1-697)

119	AspThrLeuArgLeuAlaSerSerTyrIleAlaHisIleuArgGlnIleLeuAlaAsnAsp	138
3	GACAGCTCAGGCTGGCTCCAGGTACATGCCCCACTTGAGGCAGATCTGGCTAACGAC	62
139	LysTyrGluAsnGlyTyrIleHisProValAsnLeuThrTrpProMetValAlaGly	158
63	AAATACGAGAACGGGTACATTACCCCGGTCAACCTGACGTGGGCCCTTTATGTGGCGCGG	122
159	LysProGluSerAspLeuLysGluValValThrAlaSerArgLeuCysGlyThrThrAla	178
123	AAACCCGAGATGACCTGAAGAGAGTGGTGACCGGAGCCGCTTATGTGGACCAACCGCG	182
179	Ser	179
183	TCC	185
Db		

## RESULT 15

ACH23969  
ID ACH23969 standard; cDNA: 493 BP.

ACH23969;

DT 13-OCT-2003 (first entry)

Human adult ovary cDNA #2349.

Human; ss: sequencing by hybridisation; SBH: expressed sequence tag; EST; KW: genome mapping; biodiversity; genetic disorder.

XY  
SO  
OS  
Homogametes.

XX  
PN  
US2003073623-A1

XX  
PT  
17-APR-2003

XX  
PF 30-JUN-2001 : 2001US-0091 9995

30-III-2001, 300119-00018885

XX  
DZ  
(DPMZ / ) DPMZNYC P E

PA (LABA/) LABAT I.

PA (DICK/) DICKSON M C.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

XX

XXI

PT as hybridization pro

PT antisense DNA or RNA.

Claim 1; SEQ ID NO 11181; 44pp; English.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 17:01:52 ; Search time 5521 Seconds  
(without alignments)  
10895.218 Million cell updates/sec

Title: US-09-701-674A-54  
Perfect score: 1272  
Sequence: 1 ttacggccacgactctggg.....attgtaaaaaaaaaaaaaa 1272

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.atg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sv.\*
- 13: gb.un.\*
- 14: gb.vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1272	100.0	1272	6	BD211771
2	1252.4	98.5	1254	9	AF035718 Homo sapi
3	1252.2	98.4	1257	9	AF047419 Homo sapi
4	1239	96.6	1259	6	CQ718038 Sequence
5	809.4	63.6	3231	9	BC025697 Homo sapi
6	776	61.0	1240	10	AF035717
7	774.4	60.9	1246	10	BC053525
8	764.4	60.1	1267	10	AF036945
9	761.6	59.9	1220	10	AB009453
10	759.6	59.7	1217	10	AF029753
11	752	59.1	1202	10	AF047418
12	717.2	56.4	25186	9	AL356109 Human DNA
13	581	45.7	697	6	AX333347
14	581	45.7	697	6	AX333763
15	537	42.2	537	9	CR450293 Homo sapi
16	466.2	36.7	471	6	AX333399 Sequence
17	466.2	36.7	471	6	AX333949 Sequence
18	424.6	33.4	187332	2	AC101292
19	406.2	31.9	487	9	HSPAL5B2

C	20	405.2	31.9	241544	2	AC115183
	21	387	30.4	1164	5	EX932800
	22	354.8	27.9	513	10	AF061752
	23	337.4	26.5	965	5	BC073597
	24	337.4	26.5	1059	5	AY660871
C	25	249	19.6	250	11	G15194
	26	246.8	19.4	271	6	AX896472
	27	246.8	19.4	271	6	BD032005
	28	215	16.9	124070	5	EX530074
C	29	215	16.9	231207	2	CR293516
	30	213	16.7	240	6	AR270465
C	31	212	16.7	189465	2	CR318665
	32	211.6	16.6	621	9	BT007169
	33	211.6	16.6	621	12	BT008287
	34	211.6	16.6	1917	6	CQ717187
	35	211.6	16.6	2031	9	BC066313
	36	211.6	16.6	2040	9	BC067827
	37	210	16.5	1716	9	AF087036
	38	208.6	16.4	2382	6	AX281674
	39	198	15.6	1893	9	AF060154
	40	189.6	14.9	606	10	AF108216
	41	189.6	14.9	1718	10	AF087035
C	42	171.4	13.5	170668	9	AC104012
C	43	171.4	13.5	136931	9	AC022867
C	44	160.6	12.6	225399	2	AC113255
C	45	157.4	12.4	163698	10	AC115876

ALIGNMENTS

RESULT 1	BD211771	1272 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD211771	Proteins regulating gene expression.			
DEFINITION	BD211771				
ACCESSION	BD211771.1	GI:33021541			
VERSION	JP 2002517246-A/23.				
KEYWORDS	JP 2002517246-A/23.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Lal, P., Yue, H., Tang, Y.T., Hillman, J.L., Bandman, O., Corley, N.C., Guegler, K.J., Gorgone, G.A., Baughn, M.R., Patterson, C. and Lu, D.A.M.				
TITLE	Proteins regulating gene expression				
JOURNAL	Patent: JP 2002517246-A 23 18-JUN-2002;				
COMMENT	INCYTE PHARMACEUTICALS INC				
	OS Homo sapiens (human)				
	PN JP 2002517246-A/23				
	PD 18-JUN-2002				
	PF 11-JUN-1999 JP 200053586				
	PR 12-JUN-1998 US 60/089029,29-JUL-1998 US 60/094575 PR				
	14-OCT-1998 US 60/104624				
	PI PREETI LAL,HENRY YUE,Y TOM TANG,JENNIFER L HILLMAN,OLGA PI BANDMAN.				
	PI NEIL C CORLEY,KARL J GUEGLER,GINA A GORGONE,MARIAH R BAUGHN, PI CHANDRA PATTERSON,DYUNG AINA M LU				
	PC C12N15/00,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61P15/00, PC A61E25/00,				
	PC A61P35/00,A61P37/02,A61P43/00,C07K14/47,C07K16/18,C12N15/00, CC A61K37/02				
	CC Incyte Clone No.: 3340296				
	PH Key Location/Qualifiers				
	FT 1..1272				
	Location/Qualifiers				
	1..1272				
	/organism="Homo sapiens (human)"/				
	/organism="Homo sapiens"/				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				

FEATURES

ORIGIN

Query Match 100.0%; Score 1272; DB 6; Length 1272;  
 Best Local Similarity 100.0%; Pred. No. 7e-249;  
 Matches 1272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTACGGCCACGACTCTGGAGTGGGGAACAGAGAGCGGTTCTCTGCTCCAGAGTC 60  
 DB 1 TCTACGGCCACGACTCTGGAGTGGGGAACAGAGAGCGGTTCTCTGCTCCAGAGTC 60

QY 61 CTCGGGGTTCTCTCTCACAACTCTGCGAAGGGAAAGGGTTGTGAGACCCACAGAGACC 120  
 DB 61 CTCGGGGTTCTCTCTCACAACTCTGCGAAGGGAAAGGGTTGTGAGACCCACAGAGACC 120

QY 121 CAACTCAGTCCACAGAGAGTGGCTGGCCACACTCGGAGAGCGCTCTGGTTTCAGG 180  
 DB 121 CAACTCAGTCCACAGAGAGTGGCTGGCCACACTCGGAGAGCGCTCTGGTTTCAGG 180

QY 181 GTCTCTCTGCTCTCTCTCACCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 DB 181 GTCTCTCTGCTCTCTCTCTCACCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 240

QY 241 CTCTCTCTCTGCT 300  
 DB 241 CTCTCTCTCTGCT 300

QY 301 CTTCAAGAGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTTCGAACAAGAAATT 360  
 DB 301 CTTCAAGAGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTTCGAACAAGAAATT 360

QY 361 GTGACTTCCAAACAGAGACCGAGGAGAGTCCAACTGCGAGAATGGTCTCTCCAGAG 420  
 DB 361 GTGACTTCCAAACAGAGACCGAGGAGAGTCCAACTGCGAGAATGGTCTCTCCAGAG 420

QY 421 GCGCGGGCGCTCTGCGAGAGAGAGAGGCGCCACCAAGAGAGAGCGCTCTGAGCGGG 480  
 DB 421 GCGCGGGCGCTCTGCGAGAGAGAGAGGCGCCACCAAGAGAGAGCGCTCTGAGCGGG 480

QY 481 GTCAGCAGAGAGGGAGAGAGTCCAGCGCAAGCGCCCAAGCGAGAGCGGCGCCGCG 540  
 DB 481 GTCAGCAGAGAGGGAGAGAGTCCAGCGCAAGCGCCCAAGCGAGAGCGGCGCCGCG 540

QY 541 ATCGAGTGTGAGAGAGCGCTCTCCAGACTCAGAGACCAAGCGCTGGTGGTGGCGCC 600  
 DB 541 ATCGAGTGTGAGAGAGCGCTCTCCAGACTCAGAGACCAAGCGCTGGTGGTGGCGCC 600

QY 601 GACACCAAGTCTCCAGTGGACAGCTCAGGCTGGCTCCAGCTACATCGCCCACTTG 660  
 DB 601 GACACCAAGTCTCCAGTGGACAGCTCAGGCTGGCTCCAGCTACATCGCCCACTTG 660

QY 661 AGCAGATCTGCTAACGACAATACGAGAGCGGTAATTCACCCGGTCAACTGACG 720  
 DB 661 AGCAGATCTGCTAACGACAATACGAGAGCGGTAATTCACCCGGTCAACTGACG 720

QY 721 TGCGCCCTTATGTGCGCGGAAACCGAGAGTACCTGAAAGAGTGTGACCGCGAGC 780  
 DB 721 TGCGCCCTTATGTGCGCGGAAACCGAGAGTACCTGAAAGAGTGTGACCGCGAGC 780

QY 781 CGCTTATGTGGAACCAACCGCTCTCTGCTCCAGTGGAGTGGAGTGGGAAAGCGCGCTC 840  
 DB 781 CGCTTATGTGGAACCAACCGCTCTCTGCTCCAGTGGAGTGGAGTGGGAAAGCGCGCTC 840

QY 841 CCGGGGGAGCGGGCGCGGAGAGCGCCCTGCGCTCAGTGTCTCTCTCTCTCTCTCTCT 900  
 DB 841 CCGGGGGAGCGGGCGCGGAGAGCGCCCTGCGCTCAGTGTCTCTCTCTCTCTCTCTCT 900

QY 901 CCCTCGCAATGTCT 960  
 DB 901 CCCTCGCAATGTCT 960

QY 961 ATTCTGTTTCCAAACAGAGAGATCAATTGTACTTCAAGATTCCTCATCTATTAACTT 1020  
 DB 961 ATTCTGTTTCCAAACAGAGAGATCAATTGTACTTCAAGATTCCTCATCTATTAACTT 1020

QY 1021 TATTAACCTTCTACCGTGAATGACTCTGCAAGCGCTTCTGCTCGAAGTCAATATGTAATT 1080

Db 1021 TATTAACCTTCTACCGTGAATGACTCTGCTCGAAGCTTCTGCTCGAAGTCAATATGTAATT 1080  
 QY 1081 ATAAATATAAATAGATAAGAGCGCTATCAATGTATCTTTTGTACAAATATGTTGTAAT 1140  
 Db 1081 ATAAATATAAATAGATAAGAGCGCTATCAATGTATCTTTTGTACAAATATGTTGTAAT 1140  
 QY 1141 GTAGATCATAGGATAGCTGCTTGTGACGTACATTTATAAGTAATTCATTAAAGATA 1200  
 Db 1141 GTAGATCATAGGATAGCTGCTTGTGACGTACATTTATAAGTAATTCATTAAAGATA 1200  
 QY 1201 TATATTTTTCACAAAGTTTGTCTACTTTTGAAATAAATCTTTCTTTATATTCCTAA 1260  
 Db 1201 TATATTTTTCACAAAGTTTGTCTACTTTTGAAATAAATCTTTCTTTATATTCCTAA 1260  
 QY 1261 AAAAAAAAAA 1272  
 Db 1261 AAAAAAAAAA 1272

RESULT 2  
 AF035718 1254 bp mRNA linear PRI 05-OCT-1998  
 LOCUS Homo sapiens mesoderm-specific basic-helix-loop-helix protein  
 DEFINITION (POD1) mRNA, complete cds.  
 ACCESSION AF035718  
 VERSION AF035718.1 GI:2745886  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutharia; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1254)  
 Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.  
 Pod-1, a mesoderm-specific basic-helix-loop-helix protein expressed  
 in mesenchymal and glomerular epithelial cells in the developing  
 kidney  
 Mech. Dev. 71 (1-2), 37-48 (1998)  
 JOURNAL MEDLINE 98175875  
 PUBMED 9507058  
 REFERENCE 2 (bases 1 to 1254)  
 Quaggin, S.E., Vanden Heuvel, G.B. and Igarashi, P.  
 Direct Submission  
 TITLE Submitted (24-NOV-1997) Internal Medicine, Yale University, 333  
 Cedar Street, New Haven, CT 06520-8029, USA  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source 1..1254  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 1..1254  
 /gene="POD1"  
 261..800  
 /gene="POD1"  
 /note="Pod-1"  
 /codon\_start=1  
 /product="mesoderm-specific basic-helix-loop-helix  
 protein"  
 /protein\_id="AAC62514.1"  
 /db\_xref="GI:2745887"  
 /translations="MSTGSLSDVEDLQEVLEMLCDGLKVDKSNKEFVTSNESTEESNCC  
 ENSPKRGKGLKRRRAPTKKSPLSGVQEGKQVQRNANARARARMEVLKAFRL  
 KTLPLVPPDITLKLKLDITLKLASSYTAHLRIQLANDKYENGYIHPVNLTPWPFVAKP  
 ESOLKEVVTASRLCGTTAS"  
 492..653  
 /gene="POD1"  
 /note="encodes basic-helix-loop-helix domain"

misc\_feature 492..653  
 /gene="POD1"  
 /note="encodes basic-helix-loop-helix domain"

ORIGIN  
 Query Match 98.5%; Score 1252.4; DB 9; Length 1254;  
 Best Local Similarity 99.9%; Pred. No. 6.9e-245;  
 Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



QY 8 CCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGAGAGTCCCTCGGG 67  
 Db 1 CCACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGAGAGTCCCTCGGG 60  
 QY 68 TTCCTTCTCACAACCTCTGCGAGGGGAAAGGTTGTGACACCCACACAGCCCACTCC 127  
 Db 61 TTCCTTCTCACAACCTCTGCGAGGGGAAAGGTTGTGACACCCACACAGCCCACTCC 120  
 QY 128 AGCTCCAGCAGAGAGTGGTGGCCACACTCGGAGGGCTCTTGTTTCAGGGTCTCTC 187  
 Db 121 AGCTCCAGCAGAGAGTGGTGGCCACACTCGGAGGGCTCTTGTTTCAGGGTCTCTC 180  
 QY 188 TGCTCTCTCTCACCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 247  
 Db 181 TGCTCTCTCTCACCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 240  
 QY 248 CTTCTGCTCACTCCCCCAACATGCCACGGGTTCCTCAGCATGTGAGAGCACTTCAAG 307  
 Db 241 CTTCTGCTCACTCCCCCAACATGCCACGGGTTCCTCAGCATGTGAGAGCACTTCAAG 300  
 QY 308 AGTGTGAGATGTTGGAATGTGAGGGTGTGAATGGAATTCGAACAGGAATTTGTGACTT 367  
 Db 301 AGTGTGAGATGTTGGAATGTGAGGGTGTGAATGGAATTCGAACAGGAATTTGTGACTT 360  
 QY 368 CCAACGAGACACCGAGGAGCTCCCACTGCGAGATGGGTCTCCCGAGAGGGCCGG 427  
 Db 361 CCAACGAGACACCGAGGAGCTCCCACTGCGAGATGGGTCTCCCGAGAGGGCCGG 420  
 QY 428 GCGGCTGGGCAAGAGAGGAGGCGCCCAACAGAGAGGCGCCCTGAGCGGGTCAAGC 487  
 Db 421 GCGGCTGGGCAAGAGAGGAGGCGCCCAACAGAGAGGCGCCCTGAGCGGGTCAAGC 480  
 QY 488 AGAGGGGAGCAGGTCCAGCGCAACCGCGCAACCGCGAGAGGGCGCCCATGCGAG 547  
 Db 481 AGAGGGGAGCAGGTCCAGCGCAACCGCGCAACCGCGAGAGGGCGCCCATGCGAG 540  
 QY 548 TGCTGACGAGGCTTCTCCAGACTCAAGACCACTGCGGTCGCGGTCGCGGCGGACCA 607  
 Db 541 TGCTGACGAGGCTTCTCCAGACTCAAGACCACTGCGGTCGCGGTCGCGGCGGACCA 600  
 QY 608 AGCTCTCAAGCTGAGCAGCTCAGGCTGGGTCAGCTACATCGCCCACTTGAAGCAGA 667  
 Db 601 AGCTCTCAAGCTGAGCAGCTCAGGCTGGGTCAGCTACATCGCCCACTTGAAGCAGA 660  
 QY 668 TCTGTGCTAACGACAAATACGAGACGGGTACATTCACCGGTCAACCTGAGTGGCCCT 727  
 Db 661 TCTGTGCTAACGACAAATACGAGACGGGTACATTCACCGGTCAACCTGAGTGGCCCT 720  
 QY 728 TTATGTGGCCGGAAACCGAGAGTGACTGAAGAAGTGTGACCGGAGCCGCTTAT 787  
 Db 721 TTATGTGGCCGGAAACCGAGAGTGACTGAAGAAGTGTGACCGGAGCCGCTTAT 780  
 QY 788 GTGGAACCAACCGGCTCTGACCTTGGAGTGGAGTCTGGGAAGCGCGCTCCCGGGG 847  
 Db 781 GTGGAACCAACCGGCTCTGACCTTGGAGTGGAGTCTGGGAAGCGCGCTCCCGGGG 840  
 QY 848 GAGCGGCGCCCGGAAAGGCGACCCCTGCTCTGCTCAGTGTCTCTCTCTCTCTCTCTCTC 907  
 Db 841 GAGCGGCGCCCGGAAAGGCGACCCCTGCTCTGCTCAGTGTCTCTCTCTCTCTCTCTCTC 900  
 QY 908 CAATGTCT 967  
 Db 901 CAATGTCT 960  
 QY 968 TCCAAACCAAGAGAGATCAATTTGACTTACAAAGATTCCTCATCTATTAACTTTTAAAC 1027  
 Db 961 TCCAAACCAAGAGAGATCAATTTGACTTACAAAGATTCCTCATCTATTAACTTTTAAAC 1020  
 QY 1028 TTCTACCGTAATGACTCTGCAAGCCTTGTGCTGCAAGTGTGCAATGTAATTAATAATA 1087  
 Db 1021 TTCTACCGTAATGACTCTGCAAGCCTTGTGCTGCAAGTGTGCAATGTAATTAATAATA 1080

QY 1088 TATAAATAGATAAGAGCCTATCAATGTATCTTTTGTACATATATTTGTAAATGTAGATC 1147  
 Db 1081 TATAAATAGATAAGAGCCTATCAATGTATCTTTTGTACATATATTTGTAAATGTAGATC 1140  
 QY 1148 ATAGATAGCTGACTTTTGACAGTCCACATTTATAAAGTAATTCACCTTAAAGATATATTT 1207  
 Db 1141 ATAGATAGCTGACTTTTGACAGTCCACATTTATAAAGTAATTCACCTTAAAGATATATTT 1200  
 QY 1208 TTTTCAACAAGTTTCTACTTTTGAATAAATAATCTTTCTTTATATGTCTAA 1261  
 Db 1201 TTTTCAACAAGTTTCTACTTTTGAATAAATAATCTTTCTTTATATGTCTAA 1254

RESULT 3  
 AF047419 1257 bp mRNA linear PRI 06-OCT-1998  
 LOCUS Homo sapiens epicardin mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF047419  
 VERSION AF047419.1 GI:2921852  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1257)  
 AUTHORS Robb,L., Mifsud,L., Hartley,L., Biben,C., Copeland,N.G.,  
 Gilbert,D.J., Jenkins,N.A. and Harvey,R.P.  
 TITLE epicardin: A novel basic helix-loop-helix transcription factor gene  
 of developing lung, gut, kidney, and gonads  
 JOURNAL Dev. Dyn. 213 (1), 105-113 (1998)  
 MEDLINE 98401941  
 PUBMED 9733105  
 REFERENCE 2 (bases 1 to 1257)  
 AUTHORS Robb,L. and Mifsud,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza  
 Hall Institute of Medical Research, PO Royal Melbourne Hospital,  
 Melbourne, VIC 3050, Australia  
 FEATURES  
 Location/Qualifiers  
 1..1257  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 260..799  
 /function="basic helix-loop-helix transcription factor"  
 /note="expressed in embryonic epicardium and  
 mesenchyme-derived tissues of lung, gut, kidney, and  
 gonad"  
 /codon\_start=1  
 /product="epicardin"  
 /protein\_id="AAC62532.1"  
 /db\_xref="GI:2921853"  
 /translation="MSTGSLSDVEDLQEVLEKCDGLKMDNSKEFVTSNBSSTSSNC  
 ENSPQKRGGLGKRRKAPTKPSLGSVQEGKQVQNTANARERAMRVLKAFPSRL  
 KTLTPWPPPTKLKLDLTLASSYIAHLRQILANDKIENGYIHPVNLITWFFWVAGRP  
 ESDLKEVVTYATSLCGTTAS"

CDS  
 1..1257  
 /query Match 98.4%; Score 1252.2; DB 9; Length 1257;  
 Best Local Similarity 99.8%; Pred. No. 7.6e-245;  
 Matches 1254; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

ORIGIN  
 Query Match 98.4%; Score 1252.2; DB 9; Length 1257;  
 Best Local Similarity 99.8%; Pred. No. 7.6e-245;  
 Matches 1254; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 CACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGACAGGTCCTCGGGGT 68  
 Db 1 CACGACTCTGGAGTGGGAAACAGAGAGCCGGTTCCTCTGCTGACAGGTCCTCGGGGT 60  
 QY 69 TCCTTCTCACAACCTCTCGAGAGGGAAGGTTGTGAGACCCACACAGACCCCACTCCA 128  
 Db 61 TCCTTCTCACAACCTCTCGAGAGGGAAGGTTGTGAGACCCACACAGACCCCACTCCA 120  
 QY 129 GCTCCACGAGAGGCTGCTGCGCCACACTCGGAGGCGCTCTTGGTTTCAGGGTCTCTCT 188

*Sept, bad date*

Db 121 GCTCCAGCAGGAGGTGGCTGGCGCAATTCCGAGGCTCTTGGTTTCAGGGTCTCTCT 180  
 Qy 189 GTCTCTCTCAACCTTCTCTGCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 248  
 Db 181 GTCTCTCTCAACCTTCTCTGCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 240  
 Qy 249 CTGCTCACTCCCAACATGTCACCGCTCCCTCAGCGATGTGAGACCTTCAAGA 308  
 Db 241 CTGCTCACTCCCAACATGTCACCGCTCCCTCAGCGATGTGAGACCTTCAAGA 300  
 Qy 309 GTGGAGATGTTGGAAATGTACGGGTTGAAATGGATTGCAACAGGAATTTGTACTTC 368  
 Db 301 GTGGAGATGTTGGAAATGTACGGGTTGAAATGGATTGCAACAGGAATTTGTACTTC 360  
 Qy 369 CAAGGAGACACGAGGAGCTCCAACTGCGAGATGGTCTCCCAAGAGGGCGGG 428  
 Db 361 CAAGGAGACACGAGGAGCTCCAACTGCGAGATGGTCTCCCAAGAGGGCGGG 420  
 Qy 429 CGGCTGGCAAGAGGAGGAGCGCCACCAAGAGAGCCCTCAGCGGGTCAAGCA 488  
 Db 421 CGGCTGGCAAGAGGAGGAGCGCCACCAAGAGAGCCCTCAGCGGGTCAAGCA 480  
 Qy 489 GAGGGGAGCAGGTTCAGCGCAACCGCGCAACGCGGAGAGCGGGCCGCGATCGGAT 548  
 Db 481 GAGGGGAGCAGGTTCAGCGCAACCGCGCAACGCGGAGAGCGGGCCGCGATCGGAT 540  
 Qy 549 GCTGAGCAAGGCTTCTCAGACTCAAGACCACTGCGCTGGTGGCCCGCCCGACCA 608  
 Db 541 GCTGAGCAAGGCTTCTCAGACTCAAGACCACTGCGCTGGTGGCCCGCCCGACCA 600  
 Qy 609 GCTCTCAAGCTGGACAGCTCAGGTGGCTCCAGCTACATCGCCCACTTGAAGCAGAT 668  
 Db 601 GCTCTCAAGCTGGACAGCTCAGGTGGCTCCAGCTACATCGCCCACTTGAAGCAGAT 660  
 Qy 669 CTGGGCTAAGCAATACAGAAAGGTTACATTACCCGCTCACTGACCTGACCTGCGCTT 728  
 Db 661 CTGGGCTAAGCAATACAGAAAGGTTACATTACCCGCTCACTGACCTGACCTGCGCTT 720  
 Qy 729 TATGGTGGCGGAAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCTTATG 788  
 Db 721 TATGGTGGCGGAAACCCGAGAGTGACCTGAAAGAGTGGTGACCGGAGCGCTTATG 780  
 Qy 789 TGAACCAACGCTCTGACCTTGGAGTGCGAGTGCGAGTGGGAAGCGGCTCCCGGGGG 848  
 Db 781 TGAACCAACGCTCTGACCTTGGAGTGCGAGTGCGAGTGGGAAGCGGCTCCCGGGGG 840  
 Qy 849 AGCGGGCCCGGAAAGCGACCCCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 908  
 Db 841 AGCGGGCCCGGAAAGCGACCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
 Qy 909 AATGCT 968  
 Db 901 AATGCT 960  
 Qy 969 CCAAAACAGAGGAGATCAATTGACTTCAAAAGATCCCATCTATTAACTTTTAACT 1028  
 Db 961 CCAAAACAGAGGAGATCAATTGACTTCAAAAGATCCCATCTATTAACTTTTAACT 1020  
 Qy 1029 TCTACCGTGAATGACTCTGCAAGCTTGTGCTCAAGTGCATATGTAATTTATAAATAT 1088  
 Db 1021 TCTACCGTGAATGACTCTGCAAGCTTGTGCTCAAGTGCATATGTAATTTATAAATAT 1080  
 Qy 1089 ATAAATAGATAAGAGCTTCAATGATCTTTTGTACATATGTTGTAATAATGTAGATCA 1148  
 Db 1081 ATAAATAGATAAGAGCTTCAATGATCTTTTGTACATATGTTGTAATAATGTAGATCA 1140  
 Qy 1149 TAGGATAGCTGACTTTGACAGTCACATTTATAAAGTAATTCATTTAAAGATATATTTT 1208  
 Db 1141 TAGGATAGCTGACTTTGACAGTCACATTTATAAAGTAATTCATTTAAAGATATATTTT 1200  
 Qy 1209 TTTCAACAAGATTTGCTACTTTTGAATAAATCTTTCTTTATATTGCTGTAATAAAA 1265  
 Db 1201 TTTCAACAAGATTTGCTACTTTTGAATAAATCTTTCTTTATATTGCTGTAATAAAA 1257

RESULT 4  
 LOCUS CQ718038 1259 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 3972 from Patent WO02068579.  
 ACCESSION CQ718038  
 VERSION CQ718038.1 GI:42278895  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Venter, C.J., Adams M.C., Li, P.W. and Myers, E.W.  
 TITLE Kits, such as nucleic acid arrays, comprising a majority of  
 human exons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 3972 06-SEP-2002;  
 PE Corporation (NY) (US)  
 FEATURES  
 source 1..1259  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Query Match 96.6%; Score 1229; DB 6; Length 1259;  
 Best Local Similarity 99.6%; Pred. No. 4.1e-240;  
 Matches 1254; Conservative 0; Mismatches 0; Indels 5; Gaps 2;  
 Qy 9 CACGACTCTGGAGTGGGAAAACAGAGAGCCGCTTCTCTGCTGCAAGATCTCTCGGGT 68  
 Db 1 CACGACTCTGGAGTGGGAAAACAGAGAGCCGCTTCTCTGCTGCAAGATCTCTCGGGT 60  
 Qy 69 TCCTTCTCACAACTCTCGAAGGGAAAGGTTGTGAGACCCCAACAGACCCCAACTCCA 128  
 Db 61 TCCTTCTCACAACTCTCGAAGGGAAAGGTTGTGAGACCCCAACAGACCCCAACTCCA 120  
 Qy 129 GCTCCAGCAGGAGTGGCTGCGCCACATTCGGAGGCTCTTGGTTTCAAGGTCTCTCT 188  
 Db 121 GCTCCAGCAGGAGTGGCTGCGCCACATTCGGAGGCTCTTGGTTTCAAGGTCTCTCT 180  
 Qy 189 GTCTCTCTCTCAACCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244  
 Db 181 GTCTCTCTCTCAACCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 Qy 245 CTCCTCTCTCTCACTCCCGCAACATGTCCACCGGCTCCCTCAGCGATGTGAGAGCTTC 304  
 Db 241 CTCCTCTCTCTCACTCCCGCAACATGTCCACCGGCTCCCTCAGCGATGTGAGAGCTTC 300  
 Qy 305 AAGAGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTGAAACAGAGATTTGTGA 364  
 Db 301 AAGAGTGGAGATGTTGGAATGTGACGGTTGAAATGGATTGAAACAGAGATTTGTGA 360  
 Qy 365 CTTCACACGAGAGCAGCAGGAGAGTCCAACTGCGAGAATGGGTCTCCCGAAGGGCC 424  
 Db 361 CTTCACACGAGAGCAGCAGGAGAGTCCAACTGCGAGAATGGGTCTCCCGAAGGGCC 420  
 Qy 425 GCGCGGCTTGGCAAGAGAGAGAGCGCCCAACAGAAAGAGCCCTCAGCGGGTCA 484  
 Db 421 GCGCGGCTTGGCAAGAGAGAGAGCGCCCAACAGAAAGAGCCCTCAGCGGGTCA 480  
 Qy 485 GCCAGAGGGAGAGCAGGTCCAGCGCAACCGCCCAACGCGCGAGAGC-GGGCCCGCATG 543  
 Db 481 GCCAGAGGGAGAGCAGGTCCAGCGCAACCGCCCAACGCGCGAGAGCGGGCCCGCATG 540  
 Qy 544 CGAGTGTGACAGAGGCTTCTCCAGATCAAGACCAACCTGCGCTGGGTGGGTCGCCGAC 603  
 Db 541 CGAGTGTGACAGAGGCTTCTCCAGATCAAGACCAACCTGCGCTGGGTGGGTCGCCGAC 600  
 Qy 604 ACCAGCTCTCAAGCTGGACAGCTCAGCTGGGCTCCAGCTACATCGCCCACTTGAGG 663

Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) 12477932	TITLE JOURNAL PUBMED REFERENCE
2 (bases 1 to 3231)	AUTHORS
Strausberg,R.	TITLE
Direct Submission	JOURNAL
Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov	COMMENT
Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL) CDNA Library Arrayed by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov Akter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.I., Masiello,C., Maskeri,B., Mastrian,S.D., McLoskey,J.C., McDowell,U., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgouon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.H. and Green,E.D.	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 49 Row: k Column: 7 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4507394. Location/Qualifiers 1..3231 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:34534 IMAGE:521575" /tissue_type="Pancreas, Spleen, adult pooled" /clone_lib="NIH MGC_120" /lab_host="DH10B" /note="Vector: pCMV-SPORT6"
1..3231	gene
/gene="TCF21"	CDS
/codon_start=1	
/product="transcription factor 21"	
/protein_id="AAH25697.1"	
/db_xref="GI:19344016"	
/db_xref="LocusID:6943"	
/db_xref="WIM:603306"	
/translation="MSTGSLSDVEDLQVEMLCEDGLKMDSKKEFVTNSSTEESNC ENGSPQKGRLGKRGKAPTKPSLSVGQKVORNAAPERARMVLAKAFRL KITFLWPDPPTKLTKLTLASSYLAHLQLANDKYNGYIHPVNLTPFWPVAGKP ESDLKEVVTVASRLCGTTAS"	
63.6%; Score 809.4; DB 9; Length 3231; Best Local Similarity 98.1%; Pred. No. 1.6e-154; Matches 819; Conservative 0; Mismatches 16; Indels 0; Gaps 0;	QUERY Match
9 CACGACTCTGGAGTGGGGAACAGAGAGCCGGTCTCTGTGCGAGAATCTCTCGGGT 68	
1 CACGACTCTGGAGTGGGGAACAGAGAGCCGGTCTCTGTGCGAGAATCTCTCGGGT 60	Db





		Query Match		60.1%; Score 774.4; DB 10; Length 1246;	
		Best Local Similarity		79.2%; Pred. No. 1.9e-147;	
		Matches 1019; Conservative		0; Mismatches 201; Indels 64; Gaps 6;	
Qy	3	TAGGGCCAGACTCTGGAGTGGGAAACAGAGAGCGGTTCCTCTGCTGCAAGAGTCTCT	62		
Db	13	TAGGGCCAGACTCTGGAGTGGGAAACAGAGACTGGTTCCTCTGCTGCGGGATCCT	72		
Qy	63	CGGGTTCCTCTTCAACACTCTCGAGAGGGGAAAGGTTGTGAGACCCACAGACCCCA	122		
Db	73	TGTGATCGTTTTCCAGCACTCTGTAAGGAAAGCTGGGGAGCGGACTGCCAGATCCCA	132		
Qy	123	ACTCCAGCTCCAGCAGAGGTGGTGGCCACACTCGGAGGCTCTTGGTTTCAGGGT	182		
Db	133	CCTCAAAACCAACACAGAGTGGCAGCTCCAAACTTTGAGACCTTTTCGTTTCGGTTT	192		
Qy	183	CTCTCTCTCTCTCTCAGCCCTCTTCTCGCTTTCTCTGTCTCTCTCTCTCTCTCTCT	242		
Db	193	CAGCGTCT	215		
Qy	243	CT	302		
Db	216	-----CCTCCACCTCTCTAAACATGTCACCTGCTCTCTCTCTCTCTCTCTCTCTCT	267		
Qy	303	TCAAGAGGTGGAGATGTTGAATGTGACGGGTGAAATGGATTGAAACAGGAATTTGT	362		
Db	268	TCAAGAGGTGGAGATGTTGAATGTGACGGGTGAAATGGATTGAAACAGGAATTTGT	327		
Qy	363	GATTCACAGAGACAGCAGAGAGAGTCTCAATCTCGAGATGGTCTCTCCCAAGAGG	422		
Db	328	AATTCACAGAGACAGCAGAGAGAGTCTCAATCTCGAGATGGTCTCTCCCAAGAGG	387		
Qy	423	CCCGCGCTGGCAGAGAGAGAGAGGCGCCACCAAGAGAGAGCCCTCAGCGGGT	482		
Db	388	TCGCGTGGCTAGGCAAGAGAGAGAGAGGCGCCACCAAGAGAGAGCCCTCAGCGGGT	447		
Qy	483	CAGCCAGGAGGGAAGCAGTCCAGCGCAACCGCGCAACCGCGAGAGCGGGCCGGAT	542		
Db	448	CAGCCAGGAGGGAAGCAGTCCAGCGCAACCGCGCAACCGCGAGAGCGGGCCGGAT	507		
Qy	543	GCAGTGTCTAGCAAGGCTCTTCCAGACTCAAGACACCGCTCCCTGGTGGCCCGCA	602		
Db	508	GCAGTGTCTAGCAAGGCTCTTCCAGACTCAAGACACCGCTCCCTGGTGGCCCGCA	567		
Qy	603	CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGCTCCAGTACATCGCCCACTTGAG	662		
Db	568	CACCAAGCTCTCAAGCTGACACGCTCAGGCTGGCTCCAGTACATCGCTCACTTAAG	627		
Qy	663	GCAGTGTCTAGCAAGGCTCTTCCAGACTCAAGACACCGCTCCCTGGTGGCCCGCA	722		
Db	628	GCAGTGTCTAGCAAGGCTCTTCCAGACTCAAGACACCGCTCCCTGGTGGCCCGCA	687		
Qy	723	GCCTTTATGTGGTGGCGGAAACCCGAGAGTGAACCTGAAAGAGTGGTCAACCGGAGCG	782		
Db	688	GCCTTTATGTGGTGGCGGAAACCCGAGAGTGAACCTGAAAGAGTGGTCAACCGGAGCG	747		
Qy	783	CTTATGTGGAACCAACCGCTCTGAGCTTGGAGGTGCGAGTCTGGGAAAGGCGGCTCC	842		
Db	748	CTTATGTGGAACCAACCGCTCTGAGCTTGGAGGTGCGAGTCTGGGAAAGGCGGCTCT	806		
Qy	843	GGGGGAGCGGGCGGGGAGGCGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	902		
Db	807	GGGGGAGCGGGCGGGGAGGCGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	866		
Qy	903	CCTCGCAATCT	962		
Db	867	CCTCGCAATCT	926		
Qy	963	TCGTTTCCAAACCAAGAGAGATCAATGTACTTACAAAGATTTCCATCTATTACTTTA	1022		
Db	927	TCGTTTCCAAACCAAGAGAGATCAATGTACTTACAAAGATTTCCATCTATTACTTTA	984		

Qy	1023	TTAACTTCTACCGTGAATGACTCTGCAAGCCTTGCTGGTCCCAAGTGC	1082	
Db	985	TTAACTTCTGCCATGAATGA--CTGCAAGCCTGGCTGGTCCAAAGTGC	1042	
Qy	1083	AAATATATA-----AATAGATAAGAGCCTATCAA--TGATCTTTTGTACAAT	1128	
Db	1043	ATATATATATATATATAATAATAGATAGAAGAGCCTATCAATGTGTATCTTTTGTACAAT	1102	
Qy	1129	ATGTTGTAAATATAGATCATAGATAGTCACTTTGACAGTCAATTTTATAAAGTAATT	1188	
Db	1103	ATGTTGTAAATATAGATCATAGATAGTCACTTTGACAGTCAATTTTATAAATAAGT	1162	
Qy	1189	CACCTTAAAGATATATATTTTTTTTCAACCAAGTTTTGCTACTTTTGAATAATAAACTTTTCT	1248	
Db	1163	CACCTGAAGAAATATATTTTTTTTCAACCAAGTTTTGCTACTTCCGAATAATAAACTTTTCT	1222	
Qy	1249	TTATATTGCTTAAAAAATAAAAAA 1272		
Db	1223	TTATATTGAAAAAATAAAAAA 1246		
RESULT 8				
AF036945 1267 bp mRNA linear ROD 15-JUN-1998				
LOCUS Mus musculus capsulin mRNA, complete cds.				
DEFINITION				
ACCESSION AF036945				
VERSION AF036945.1 GI:2695694				
KEYWORDS				
SOURCE Mus musculus (house mouse)				
ORGANISM Mus musculus				
REFERENCE				
AUTHORS Lu,J.,Richardson,J.A. and Olson,E.N.				
TITLE Capsulin: a novel bHLH transcription factor expressed in epicardial progenitors and mesenchyme of visceral organs				
JOURNAL Mech. Dev. 73 (1), 23-32 (1998)				
MEDLINE 98213628				
PUBMED 9545521				
REFERENCE				
AUTHORS Lu,J. and Olson,E.N.				
TITLE Direct Submission				
JOURNAL Submitted (04-DEC-1997) Molecular Biology and Oncology, UT Southwestern Medical Center, 6000 Harry Hines Blvd., NAB.510, Dallas, TX 75235-9148, USA				
FEATURES				
source				
1..1267				
/organism="Mus musculus"				
/mol_type="mRNA"				
/db_xref="taxon:10090"				
264..803				
/function="bHLH transcription factor"				
/codon_start=1				
/product="capsulin"				
/protein_id="AAC23537.1"				
/db_xref="GI:2695695"				
/translation="MSTGSLSDVDELDQEVMLDLSIKVDSNKEFGTSNESTBEGSNCE				
ENSPQKRGRLGKRRKAPTKPSPLSGVQEGKQVORNAARERARMLVSKAFSL				
KTLTPVPFPDITKLSDITLKLASSYIAHLRLANDKYENGYIHPVNLTFPFWVAKP				
ENDLKEVVTANRLCGTTAS"				
ORIGIN				
Query Match 60.1%; Score 764.4; DB 10; Length 1267;				
Best Local Similarity 79.2%; Pred. No. 1.9e-145;				
Matches 1009; Conservative 0; Mismatches 201; Indels 64; Gaps 6;				
Qy	3	TACGGCCAGCACTCTGGAGTGGGAAACAGAGAGCGGTTCCTCTGCTGCAAGAGTCTCT	62	
Db	44	TAGGCCAGGACTCCAGTGGTGGGAGATAGAGCACTGGTTCCCTGTGCGGGATCCCT	103	
Qy	63	CGGGTTCCTTCTCACAACTCTCGGAAGGGGAAAGGGTTGTGAGACCCACAGACCCCA	122	
Db	104	TGTGATCGTTTTCCAGCACTCTGTAAGAGGAAGAGCTGGGGAGCGGACTGCCAGATCCCA	163	

Qy	123	ACTCAGCTCCACGAGAGGTGCTGCGCCACACTCGGAGGCCTCTTGTTTCAGGGT	182
Db	164	CCTCAAAACCCAAACACACGAAAGTGCAGCTCCAAAACCTTTGAGACCCCTTCGTTTCCGTTT	223
Qy	183	CTCTCTGTCTCTCTCTCACCCTCTTCCTCGGCTTTCTGTGTCTCTCTCTCTCTCTCTCT	242
Db	224	CAGCGTCTCTCTCTCCCTCCCTC-----	246
Qy	243	CTCTCCCTGTCACACTCCCCCAACATGTCTCACGGGTCCCTCAGCGATGFGAGACCT	302
Db	247	-----CCTCCACCTCTCTTAAACATGTCTCACTGGTCCCTCAGCGATGTAGAGACCT	298
Qy	303	TCAAGAGTGGAGATGTTTGAATCTGAGGGTTCGAAATGATTCGACACAGGAATTTGT	362
Db	299	TCAAGAGTGGAGATGCTGAGTCTGACTCCTGAAAGTGGACTCCAAACAGGAGTTTGG	358
Qy	363	GACTTCCAAACGAGAGCACCGAGGAGAGCTCCAACTGCGAGAAATCGGTCTCCCCAGAGGG	422
Db	359	AACTTCCAAACGAGAGCACCGAGGAGGGTCCAACTGCGAGAAACGGGTCTCCACAGAGGG	418
Qy	423	CCGCGCGGCTCGGCAAGAGAGGAGAGGGCCCAACCAAGAGAGCCCCCTGAGCGGGT	482
Db	419	TCGCGGTGCTTAGCAAGAGAGGAGAGGGCCCACTAAGAAAGCCGCTCAGCGGGGT	478
Qy	483	CAGCCAGAGGGGAAGCAGGTCCAGCGCAACGCCGCCAAACGCGGAGAGGGGCCCGCAT	542
Db	479	CAGCCAGAGGGCAAGCAGGTCCAGCGCAACCGCGCCAAATGCTCGTGAAGGGGCGCGCAT	538
Qy	543	GCAGTCTGTCAGAGAGCCCTTCTCCAGACTCAAGACCAACCTGCGCTGGGTGCCCCCGA	602
Db	539	CGGGGTCTGAGCAAGGCCCTTCTCCAGCTCAAGACCAACCTGCGCTGGGTGCCCCCGGA	598
Qy	603	CACCAAGCTCTCCAAGCTGGACACGCTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAG	662
Db	599	CACCAAGCTCTCCAAGCTGGACACTCTCAGCTGGCGTCCAGCTACATCGCTCACTTAAG	658
Qy	663	GCAGTCTGGCTAACGACAAATACGAGAACGGGTACATTACCCGGTCAACCTGACGTG	722
Db	659	GCAGTCTGGCCCAACGACAAAGTACGAAACGGTTAATTACCCAGTCAACCTGAGCGT	718
Qy	723	GCCCTTTATGTTGGCCGGGAACCCGAGAGTGAACCTGAAAGAGTGTGTACCCGCGAGCCG	782
Db	719	GCCCTTTATGTTGGCCGGCAACACAGAGATGACCTGAGGAAGTGTGTACCCGCAACCG	778
Qy	783	CTTATGTGAACACCGCGTCTGACCTTGGAGGTGCGAGTCTGGGAAAGCGCGCTCCC	842
Db	779	CTTGTGTGAACATCAGCATCTGACCTTGGAGGTGCGAGTCTGGGAACCGTGAGCT-CT	837
Qy	843	GGGGGAGCGGGCCCGGGGAAGCGACCCCTGCGCTCAGTGTCTCTGTCTCTGTCTTCCC	902
Db	838	GGGGGAGTGGGCGGGCAAGGGTGAAGACCCCTGCTCCCTCTCTGCTGTGGTG	897
Qy	903	CCTCGAATGCTCTCTCTGTGCCACCCCGGAGAAACATTTAACAGCAGAGGAGAT	962
Db	898	CCTTGCAATGCTCTCTCTCCCTATCCCAACCCCGAAGAACACTTAACTCGACAGGAGAT	957
Qy	963	TCGTTTCCAAACAGAGAGATCAATGTACTTACAAAGATTCCTCATCTATTTAACTTTA	1022
Db	958	TCGTTTCCAAACAAAGAGACCAACTGTACTTAC--CGATTCTCATCTATTAACTTTGA	1015
Qy	1023	TTAACTTCTACCGTGAATGACTCTGCAAGCGCTTGCTGGTCAAGTCAATATGTAATAT	1082
Db	1016	TTAACTTCTGCCATGAATGA--CTGCAAGCGCTGGCTGGTCAAGTCAATATGTAATAT	1073
Qy	1083	AAATATATA-----AATFAGATAAGGCCTATCAA--TGTATCTTTTGTACAAT	1128
Db	1074	ATATATATATATATATAATATAGATAGAGAGCCTATCAATGTGTATCTTTTGTACAAT	1133
Qy	1129	ATGTTGAAAATGATGATCATATAGATAGCTGACTTTGACAGTCACATTTATAAGTAATT	1188
Db	1134	ATGTTGAAAATGATGATCATATAGATAGCTGACTTTGACAGTCACATTTATAAAATAAGT	1193

QY	1189	CAC	TTAAGATATATATATTTTCAACAAGATTTTGTCTACTTTTGAATAAATCTTTCT	1248
DB	1194	CAC	TCGAAGAAATATATATTTTCAACAGGTTTGTCTACTTCGAAATAAATCTTTCT	1253
QY	1249	TTA	TATTGCTAAAA	1262
DB	1254	TTA	TATTGAAAAA	1267
RESULT 9				
AB009453				
LOCUS	AB009453		1220 bp	mRNA linear ROD 11-DEC-1997
DEFINITION	Mus musculus mRNA for nephogonadin, complete cds.			
ACCESSION	AB009453			
VERSION	AB009453.1	GI:2696117		
KEYWORDS	nephogonadin.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 1220)			
TITLE	Tamura, M. and Nakatsuji, N.			
JOURNAL	Identification of Nephogonadin, a novel basic helix-loop-helix gene Published Only in DataBase (1997)			
REFERENCE	2 (bases 1 to 1220)			
AUTHORS	Tamura, M. and Nakatsuji, N.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-DEC-1997) Masaru Tamura, National Institute of Genetics, Mammalian Development; Yata 1111, Mishima, Shizuoka 411, Japan (E-mail:matamura@lab.nig.ac.jp, Tel:+81-559-81-6832, Fax:+81-559-81-6828)			
FEATURES	Location/Qualifiers			
source	1..1220			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/db_xref="taxon:10090"			
CDS	223..762			
	/function="transcription factor"			
	/note="basic helix-loop-helix gene"			
	/codon_start=1			
	/product="nephogonadin"			
	/protein_id="BA23883.1"			
	/db_xref="GI:2696118"			
	/translations="MSGTGLSDVEDLQEVMLDCDSLKVDSNKEFGTSNESTEEGNC ENGSPKRGGLGRKRAKTPKSLGVSQEGKQVORNAARARMVLSKAFRL KTLTPWPPTKLSKLDLRLASSYIAHLRQLANDKYENGVIHPVNLWTFWVAVGKPN ENDLKEVVFANRLCGTTAS"			
	polyA_site			
	1220			
	/note="17 A nucleotides"			
ORIGIN				
Query Match	59.9%	Score 761.6;	DB 10;	Length 1220;
Best Local Similarity	79.3%;	Pred. No. 7.1e-145;		
Matches 1005;	Conservative	0;	Mismatches 199;	Indels 64; Gaps 6;
QY	3	TAGGGCCACACTCTGGAGTGGGGAACGAGAGCGGTTCTCTCTGTCGAGAAATCCT	62	
DB	3	TAAGGCCAGACTCCAGTGTGTGGGAGATAGACACTGGTTTCCCTGTGCGGGATGCCCT	62	
QY	63	CGGGGTTCTCTTCACAACTCTCGGAAGGGGAAGGTTGTGAGACCCCAACAGACCCCA	122	
DB	63	TGTGATCGTTTTCCACGACTCTGTAAGGGAAGCTGGGGAGCGGACTGCCAGATCCCA	122	
QY	123	ACTCCAGCTCCCAAGCAGAGAGGTGGTGGCCACACTCGGGAGGCCTCTTGGTTTCAGGCT	182	
DB	123	CCTCAAAACCAACACACGAAAGTGGCAGCTCCAAACTTTTGAGACCCCTTCGTTTCGGTTT	182	
QY	183	CTCTCTGCTCTCTCTCACCCCTCTTCTCGCTTTCTCTGTCCTCTCTCTCTCTCTCT	242	
DB	183	CAGGTCCTCTCTCTCCCTCCCTC-----	205	
QY	243	CTCTCCCTCGTCCACTCCCAACATGTCACCGGCTCCCTCAGCGATGTGGAGACCT	302	







QY 372 CGAGAGCACCGAGAGAGCTCCAACTGCGAGAAATGGTCTCCCAAGAGCGCGCGG 431  
 Db 317 CGAGAGCACCGAGAGAGCTCCAACTGCGAGAAATGGTCTCCCAAGAGCGCGCGG 376  
 QY 432 CTTGGGCAAGAGAGAGAGCGCCCAACCAAGAGAGAGCGCCCTGAGCGGGTCAAGCAGA 491  
 Db 377 CTTAGGCAAGAGAGAGAGCGCCCAACCAAGAGAGAGCGCCCTGAGCGGGTCAAGCAGA 436  
 QY 492 GGGGAAGCAGGTCAGCGCAACCGCCCAACCGCGAGAGAGCGCGCCCTGAGCGAGTGT 551  
 Db 437 GGGCAAGCAGGTCAGCGCAACCGCCCAACCGCGAGAGAGCGCGCCCTGAGCGAGTGT 496  
 QY 552 GAGCAAGGCTCTTCAGAGCTCAAGACCAACCTGCGCTGGTGGTCCCCCGACACCAAGCT 611  
 Db 497 GAGCAAGGCTCTTCAGAGCTCAAGACCAACCTGCGCTGGTGGTCCCCCGACACCAAGCT 556  
 QY 612 CTCAGAGCTGGACACAGCTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAAGCAGATCCT 671  
 Db 557 CTCAGAGCTGGACACAGCTCAGGCTGGCGTCCAGCTACATCGCTCACTTGAAGCAGATCCT 616  
 QY 672 GGCTAAGCAAAATAGAGAACGGGTACATTCACCGGTCAACCTGAGCTGAGCGCTTTAT 731  
 Db 617 GGGCAAGCAGCAAGTACAGAACGGGTACATTCACCGGTCAACCTGAGCTGAGCGCTTTAT 676  
 QY 732 GGTGGCGGGAACCGAGAGTCACTGAAAGAGTGTGACCGGAGCGGCTTATGTGG 791  
 Db 677 GGTGGCGGGAACCGAGAGTCACTGAAAGAGTGTGACCGGAGCGGCTTATGTGG 736  
 QY 792 AACCAAGGCTCTGACCTTGGAGTCCGAGTCTGGGAAAGCGCGCTCCCGGGGAGC 851  
 Db 737 AACCAAGGCTCTGACCTTGGAGTCCGAGTCTGGGAAAGCGCGCTCCCGGGGAGC 795  
 QY 852 GGGCGCGGGAAGCGGACCGCTGCCCTCAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 911  
 Db 796 GGGCGCGGGAAGGAGTACGACCGCTGCCCTCAGTGTCTCTGCTGCTGCTGCTGCTGCTGCT 855  
 QY 912 GCTCCTCTCTCTCTCCACCGCGGAGACACCTTTACAAAGAGAGGAGTTCGTTTCA 971  
 Db 856 GCTCCTCTCTCTCTCCACCGCGGAGACACCTTTACAAAGAGAGGAGTTCGTTTCA 915  
 QY 972 AACCAAGGAGTCAATGTGATCAAGAGTCCCAATCTATTTAATTTAATTTAATTTAATTT 1031  
 Db 916 AACCAAGGAGTCAATGTGATCAAGAGTCCCAATCTATTTAATTTAATTTAATTTAATTT 973  
 QY 1032 ACCGTGAATCACTCTGCAAGCCTTGTGTTCCAGTCAATATGTAATTAATTAATTAATTA 1091  
 Db 974 GCCATGATGA--CTGCAAGCCTGCTGTTCCAGTCAATATGTAATTAATTAATTAATTA 1031  
 QY 1092 -----AATGATAAGAGCCTATCAA--TGATCTTTTGTACAAATATGTTGTA 1137  
 Db 1032 TTATATAAATAGATAAGAGCCTATCAATGTGTATCTTTTGTACAAATATGTTGTA 1091  
 QY 1138 AATGTAGATCATAGTACGCTTGTGAGTCAATTTAATTAATTAATTAATTAATTAATTA 1197  
 Db 1092 AATGTAGATCATAGTACGCTTGTGAGTCAATTTAATTAATTAATTAATTAATTAATTA 1151  
 QY 1198 ATATATATTTTTTCAACAAAGTTTTGTACTTTTCAAAATTAATCTTTCTTTATATG 1257  
 Db 1152 AAATATATTTTTTCAACAAAGTTTTGTACTTTCCGAAATTAATCTTTCTTTATATGA 1211  
 QY 1258 TAAAA 1263  
 Db 1212 AAAAAA 1217

RESULT 11  
 AF047418  
 LOCUS AF047418 1202 bp mRNA linear ROD 06-OCT-1998  
 DEFINITION Mus musculus epicardin mRNA, complete cds.  
 ACCESSION AF047418  
 VERSION AF047418.1 GI:2921850  
 KEYWORDS Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 1202)  
 AUTHORS Robb, L., Mifusud, L., Hartley, L., Biben, C., Copeland, N.G.,  
 Gilbert, D.J., Jenkins, N.A. and Harvey, R.P.  
 TITLE epicardin: A novel basic helix-loop-helix transcription factor gene  
 expressed in epicardium, branchial arch myoblasts, and mesenchyme  
 of developing lung, gut, kidney, and gonads  
 Dev. Dyn. 213 (1), 105-113 (1998)  
 JOURNAL 98401941  
 MEDLINE 9733105  
 PUBMED 9733105  
 REFERENCE 2 (bases 1 to 1202)  
 AUTHORS Robb, L. and Mifusud, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-1998) Haematology and Cancer, Walter and Eliza  
 Hall Institute of Medical Research, PO Royal Melbourne Hospital,  
 Melbourne, VIC 3050, Australia  
 FEATURES  
 Location/Qualifiers  
 1..1202  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Balb/c"  
 /db\_xref="taxon:10090"  
 199..738  
 /function="basic helix-loop-helix transcription factor"  
 /note="expressed in embryonic epicardium and  
 mesenchyme-derived tissues of lung, gut, kidney, and  
 gonad"  
 /codon\_start=1  
 /product="epicardin"  
 /protein\_id="AAC62531.1"  
 /db\_xref="GI:2921851"  
 /translation="MSTGSLSDVEDLQEVMLDCDSLKVDNSKEFGTSNESTEGSNC  
 ENGSPQGRGLGKRRKAPTKGSLGVQSGQVQRNANARARMLVSKAFSL  
 KTLPLWPPDTKLKSLDRLASSYTAHLRQILANDKYENGYIHPVNLTPFWVAGKP  
 ENDLKVVVTANRLCGTTAS"  
 ORIGIN  
 Query Match 59.1%; Score 752; DB 10; Length 1202;  
 Best Local Similarity 79.1%; Pred. No. 6.4e-143;  
 Matches 993; Conservative 0; Mismatches 195; Indels 64; Gaps 6;  
 QY 25 GGGAAACAGAGAGCGGTTCTCTGCTGCAGAAAGTCTCGGGGTCTCTTCTCAACAATCT 84  
 Db 1 GGGAGATAGACATGTTTCCCTGTCGGGATCCCTTGATCGTTTCCACGACTCT 60  
 QY 85 CGAAGGGGAAGGGTGTGAGACCCAAACAGACCCCAACTCCAGTCCCGAGGAGGT 144  
 Db 61 GTAAAGGGAAGCTGGGGAGCGGACTCCAGATCCCACTCAAAACCAACACAGAGT 120  
 QY 145 GGCTGGCCACACTCGGGAGGCTTTGGTTTCAAGGCTCTCTGCTCTCTCTCACCT 204  
 Db 121 GGCAGCTCCAACTTTTGAGACCTTTCGTTTCGGTTTACGGTCTCTCTCTCTCTCT 180  
 QY 205 CTTCTCGCTTCT 264  
 Db 181 C-----CCTCCACCTCTA 195  
 QY 265 AACATGTCACCGCTCCCTCAGAGATGTGGAGACCTTCAAGAGGTGGAGATTTGGAA 324  
 Db 196 AACATGTCACCTGCTCCCTCAGAGATGTGGAGACCTTCAAGAGGTGGAGATTTGGAG 255  
 QY 325 TGTACGGGTTGAAAATGGATTCGAACAAGGAATTTGTGACTTCCACAGAGACACCGAG 384  
 Db 256 TGTGACTCCCTGAAAGTGGACTCCAAAGAGTTTGGAACTTCCACAGAGACACCGAG 315  
 QY 385 GAGAGCTTCAACTCGGAGATGGGTCTCCCAAGAGGCGCGCGCTTGGGCAAGAG 444  
 Db 316 GAGGCTTCCAACTCGGAGAGCGGCTCTCCAGAGAGGTCGCGTGGCTAGGCAAGAG 375  
 QY 445 AGGAAGGCGGCCCAACAGAGAGAGCCCTCTGAGCGGGGTCAAGAGGAGGAGGAGCAGTC 504

376	AGGAAGCGCCCACTAAGAAAAGCCCGCTCAGCGGGGTGACGACGAGGGCAGCAGTGC	435
505	CAGCGCAACGCGCGCCAAACGCGAGAGCGGGGCCCGCATCGAGTGTGAGCAAGCCCTTC	564
436	CAGCGCAACGCGGCCAATGCTCGTAGCGGGCCCGCATGCGGTGCTGAGCAAGGCCCTTC	495
565	TCGAGACTCAGACCAACCCCTGCTCGGTGCGCCCGGACACCAAGCTCTCCAGCTGCAC	624
496	TCGAGGCTCAAGACCAACCCCTGCGGTGCGCCCGGACACCAAGCTCTCCAAGTGCAC	555
625	ACGCTCAGGCTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCTGGCTAAACACAAA	684
556	ATCTCAGGCTGGGTCCAGTCAATCGCTCACTTAAGGCGAGATCTGGCCAAACGACAG	615
685	TACGAGAACGGGTACATTCACCCGGTCAACCTGACCTGGCCCTTTATGTTGGCCGGGAAA	744
616	TACGAGAACGGTTACATTCACCCAGTCAACCTGACCTGGCCCTTTATGTTGGCCGGCAA	675
745	CCGAGAGTGACCTGAAGAAGTGGTGAACCGGAGCGCGCTTATGTGGAAACACACCGTCC	804
676	CCAGAGAAATGACCTGAAGGAAAGTGGTCAACCGCCAAACCGCTTGTTGGAACTACAGCATCC	735
805	TGACCTTGGAGTGGAGTCTTGGAAAAGCGCGCTCCCGGGGGAGCGGGCCCCGGGAAG	864
736	TGACCTTGGAGTGGAGTCTTGGAAAACCGTAGCT-CTGGGGCAGTGGCGCGGCAAGGG	794
865	GGGACCCCTGGCCCTCAGTGCTCTGTGCTGTGCTTCCCGCTGGCAANGTCTCTCTCTG	924
795	TGACGACCCCTGGCCCTTCCCTTCCCTCTGCTCTGCTGGTGGCTTGTCATCGCTCTCTCCCTA	854
925	TCCGACCCCGGAGAACACTTTACAAAGAGGAGATTCGTTTCCAAACACGAGGAGAT	984
855	TCCGACCCCGGAGAACACTTAACTGGGAGGAGGAGATTCGTTTCCAAACCAAGAGGAC	914
985	CAATTGTACTTACAAAGATTCCTCATCTATTATTAATCTTAACTCTACCGTGAATGACT	1044
915	CAACTGTACTTAC--CGATTCTCATCTATTTAATCTTGATTAACTCTGCCATGATGA--	970
1045	CTGCAAGCCCTTGCTGGTCCAGTCCGAATGTAATTAATAATAATA-----A	1092
971	CTGCAAGCCCTGGCTCCGAGTCCAAATGTAATTAATAATAATAATAATAATAATAATA	1030
1093	ATAGATAGAGCCCTATCAA--TGTATCTTTTGTACAAATGTTGTGAAAATGTAGATCATA	1150
1031	GATAGAGAGCCCTATCAATGTGTATCTTTGTACAAATGTTGTGAAAATGTAGATCATA	1090
1151	GGATAGCTGACTTTGACAGTCAACATTTATAAAGTAATTCACCTTAAAGATATATATTTT	1210
1091	GAATAGCTGACTTTGACAGTCAACATTTATAAATAAGTCACTCGAAGAAATATAATTTT	1150
1211	TCAAACAGATTTTGCTATCTTTGAAAATAAATCTTCTTTATATGCTGCTAAAA	1262
1151	TCAAACAGATTTTGCTATCTCCGAAAATAAATCTTCTTTATATGAAAATA	1202

RESULT 12	
AL356109	LOCUS
AL356109	DEFINITION
AL356109	25186 bp DNA linear PRI 23-JAN-2001
AL356109	HUMAN DNA sequence from clone RP11-373A10 on chromosome 6. Contains
AL356109	ESTs, STSS, GSSs and CPG islands. Contains the TCF21 gene for
AL356109	transcription factor 21, complete sequence.
AL356109	ACCESSION
AL356109.8	VERSION
GI:1043434	KEYWORDS
HTG: Cpg island; TCF21; transcription factor 21.	SOURCE
Homo sapiens (human)	ORGANISM
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 25186)	
Wallis.J	Direct Submission
Submitted (15-DEC-2000)	Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerv@sanger.ac.uk	Clone
JOURNAL	
TITLE	
AUTHORS	
REFERENCE	

requests: clonerequest@sanger.ac.uk

On Oct 1, 2000 this sequence version replaced gi:10186139.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: ER., EMBL, Sw., SWISSPROT, Tr., TREMBL, Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

IMPORTANT: This sequence is not the entire insert of clone RP11-373A10 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP1-73H22 is at 25087 in this sequence. The true right end of clone RP4-662A9 is at 100 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

RP11-373A10 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES	source
repeat.	repeat.
repeat.	repeat.
repeat.	repeat.
repeat.	repeat.
misc_f	misc_f
repeat.	repeat.
repeat.	repeat.
repeat.	repeat.
repeat.	repeat.
misc_f	misc_f
repeat.	repeat.
repeat.	repeat.
repeat.	repeat.
repeat.	repeat.
gene	gene
mRNA	mRNA

```
/product="ba373a10.1 (transcription factor 21)"
/notes="match: cDNAs: Em:AB009453 Em:AF036945 Em:AF029753
Em:AF035717 Em:AF047418 Em:AF035718 Em:AF047419
Em:AF087036 Em:AF060154 Em:AF108216 Em:M99367 Em:AF061752
Em:Y11255 Em:L35922 Em:AF109012 Em:M86923 Em:Y09596
Em:AB007959 Em:U18658
match: ESTs: Em:AW313610 Em:W73859 Em:AA880760 Em:A1927658
Em:R22321 Em:AW357120 Em:AW432075 Em:AW190108 Em:AW025547
Em:AW242648 Em:AW548204 Em:A1971939 Em:W73825 Em:R65626
Em:W08124 Em:AW558915 Em:AW086478 Em:AA154683 Em:AA028184
Em:AA027032 Em:AA335365 Em:A1478874 Em:A1578558 Em:N95204
Em:R63028 Em:AV374791 Em:W21552 Em:AA028092 Em:N66932
Em:AA358824 Em:A1712521 Em:A1711406 Em:C21328 Em:A1185851
Em:H03852 Em:AA834021 Em:A1721064 Em:AV365142 Em:AV366354
Em:AA027107 Em:AV311062 Em:AV365865 Em:AA341736
Em:A1336106 Em:AA905950 Em:AV351072 Em:A1178188
Em:AA219936 Em:AW197130 Em:HS1772 Em:AA340993 Em:AV351557
Em:D45696 Em:H95174"
/evidence=not experimental
13096. .13165
/notes="35 copies 2 mer to 85% conserved"
join(13182, 13531, 15497, .15586)
/gene="TCF21"
/notes="match: proteins: Tr:Q62523 Tr:O88940 Tr:P79782
Tr:O75946 Tr:Q43545 Tr:Q64124 Tr:O73615 Sw:Q13562
Tr:Q91418 Tr:Q61100 Tr:Q922E5 Tr:Q9YH81 Tr:Q9YH82
Tr:Q90690 Sw:P79920 Sw:Q60867 Sw:P24899 Tr:Q9W6C6
Tr:Q9PT60 Sw:Q60430 Tr:Q9PTE4 Sw:Q64289 Sw:P79765
Tr:O60682 Tr:O70483 Tr:Q9YHA3 Tr:O95300 Tr:Q61246
Tr:Q98939 Tr:Q9YH69 Tr:P797831 Sw:Q91616 Sw:P79766
Tr:P797833"
/codon_start=1
/evidence=not experimental
/product="ba373a10.1 (transcription factor 21)"
/protein_id="CAC26822.1"
/db_xref="GI:12339605"
/db_xref="GOA:O43680"
/db_xref="Swiss-Prot:O43680"
/translations="METGLSDVDELDQVEVMECDGLKSDNKEFYTSNTEESSNC
ENGSPKQSGGLGKRKAPTKSLQSGVQEGQVKNANARERARMEVLGKAFSRL
KTLPLWPPDTKLSDTLRLASSYIAHLRQLANDKYENGVIHPVNLTPWFWAGKP
ESDLKEVVTASRLCGITAS"
13286. .13864
/gene="TCF21"
/notes="Cpg island"
/evidence=not experimental
13326. .13636
/gene="TCF21"
/notes="match: GSS: Em:AL212694 Em:AL254874"
complement(13513, .13636)
/notes="match: GSS: Em:AL214658"
15337. .15744
/gene="TCF21"
/notes="Cpg island"
/evidence=not experimental
complement(15683, .16091)
/notes="match: GSS: Em:AQ031468"
complement(15788, .16037)
/notes="match: STS: Em:G15194"
16015. .16020
/gene="TCF21"
16041
/gene="TCF21"
17643. .17736
/notes="MIR repeat: matches 49. .156 of consensus"
18016. .18073
/notes="29 copies 2 mer ct 72% conserved"
18787. .18822
/notes="18 copies 2 mer ct 97% conserved"
19018. .19113
/notes="MER5A repeat: matches 1. .121 of consensus"
complement(19104, .19320)
/notes="match: STS: Em:G21594"

misc_feature
1 TCTACGGCCACGACTCTGGGAGTGGGAAACAGAGAGCCGGTCTCTCTGTCGAGAGTC 60
12915 TCTACGGCCACGACTCTGGGAGTGGGAAACAGAGAGCCGGTCTCTCTGTCGAGAGTC 12974
QY 61 CTCGGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGGTTGTGAGACCCCAACAGAGCC 120
12975 CTCGGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGGTTGTGAGACCCCAACAGAGCC 13034
QY 121 CAATCCAGCTCCAGAGAGAGTGGCTGCCCACTCGGAGAGCCCTCTTGGTTTCAGG 180
13035 CAATCCAGCTCCAGAGAGAGTGGCTGCCCACTCGGAGAGCCCTCTTGGTTTCAGG 13094
QY 181 GTCTCTGTCTCTCTCACCCTCTTCTCGCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
13095 GTCTCTGTCTCTCTCTCACCCTCTTCTCGCTTTCTCTCTCTCTCTCTCTCTCTCTCT 13154
QY 241 CTCTCTCCTCGTCCACTCCCCCAAAATGTCACCGGCTCCCTCAGCGATGTGGAGAC 300
13155 CTCTCTCCTCGTCCACTCCCCCAAAATGTCACCGGCTCCCTCAGCGATGTGGAGAC 13214
QY 301 CTTCAAGAGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTGGAACAAGGAATTT 360
13215 CTTCAAGAGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTGGAACAAGGAATTT 13274
QY 361 GTGACTTCCAAACGAGAGACCCAGAGAGTCCAACTCGGAGATGGGTCTCCCCAGAG 420
13275 GTGACTTCCAAACGAGAGACCCAGAGAGTCCAACTCGGAGATGGGTCTCCCCAGAG 13334
QY 421 GGCCGCGCGGCTGGGCAAGAGAGGAGGAGGCGGCCCAACAGAGAGCCCTCTGAGCGG 480
13335 GGCCGCGCGGCTGGGCAAGAGAGGAGGAGGCGGCCCAACAGAGAGCCCTCTGAGCGG 13394
/notes="match: STS: Em:G21594"

ORIGIN
Query Match 56.4%; Score 717.2; DB 9; Length 25186;
Best Local Similarity 99.6%; Pred. No. 1.4e-135;
Matches 719; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTACGGCCACGACTCTGGGAGTGGGAAACAGAGAGCCGGTCTCTCTGTCGAGAGTC 60
Db 12915 TCTACGGCCACGACTCTGGGAGTGGGAAACAGAGAGCCGGTCTCTCTGTCGAGAGTC 12974
QY 61 CTCGGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGGTTGTGAGACCCCAACAGAGCC 120
Db 12975 CTCGGGGTTCCTTCTCACAACCTCTCGAAGGGGAAAGGGTTGTGAGACCCCAACAGAGCC 13034
QY 121 CAATCCAGCTCCAGAGAGAGTGGCTGCCCACTCGGAGAGCCCTCTTGGTTTCAGG 180
Db 13035 CAATCCAGCTCCAGAGAGAGTGGCTGCCCACTCGGAGAGCCCTCTTGGTTTCAGG 13094
QY 181 GTCTCTGTCTCTCTCACCCTCTTCTCGCTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 13095 GTCTCTGTCTCTCTCTCACCCTCTTCTCGCTTTCTCTCTCTCTCTCTCTCTCTCTCT 13154
QY 241 CTCTCTCCTCGTCCACTCCCCCAAAATGTCACCGGCTCCCTCAGCGATGTGGAGAC 300
Db 13155 CTCTCTCCTCGTCCACTCCCCCAAAATGTCACCGGCTCCCTCAGCGATGTGGAGAC 13214
QY 301 CTTCAAGAGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTGGAACAAGGAATTT 360
Db 13215 CTTCAAGAGTGGAGATGTTGGAATGTGACGGGTTGAAATGGATTGGAACAAGGAATTT 13274
QY 361 GTGACTTCCAAACGAGAGACCCAGAGAGTCCAACTCGGAGATGGGTCTCCCCAGAG 420
Db 13275 GTGACTTCCAAACGAGAGACCCAGAGAGTCCAACTCGGAGATGGGTCTCCCCAGAG 13334
QY 421 GGCCGCGCGGCTGGGCAAGAGAGGAGGAGGCGGCCCAACAGAGAGCCCTCTGAGCGG 480
Db 13335 GGCCGCGCGGCTGGGCAAGAGAGGAGGAGGCGGCCCAACAGAGAGCCCTCTGAGCGG 13394
/notes="match: STS: Em:G21594"
```

```

QY 481 GTGAGCAGAGGGGAAGCAGGTCCAGCGCAACGCCGCCCAACGCCGAGAGCGGGCCCGC 540
Db 13395 GTGAGCAGAGGGGAAGCAGGTCCAGCGCAACGCCGCCCAACGCCGAGAGCGGGCCCGC 13454
QY 541 ATGGAGTGTGAGCAAGGCGCTTCCAGACTCAAGACCAACCCCTGCCCTGGGTGCCCGCC 600
Db 13455 ATGGAGTGTGAGCAAGGCGCTTCCAGACTCAAGACCAACCCCTGCCCTGGGTGCCCGCC 13514
QY 601 GACACCAAGCTCTCCAAAGTGGACAGCTCAGCGCTGGCGTCCAGCTACATCGCCCACTTG 660
Db 13515 GACACCAAGCTCTCCAAAGTGGACAGCTCAGCGCTGGCGTCCAGCTACATCGCCCACTTG 13574
QY 661 AGGAGATCTCTGCTACGACCAATACGAGACGGGTACATTCACCGCTCAACCTGACG 720
Db 13575 AGGAGATCTCTGCTACGACCAATACGAGACGGGTACATTCACCGCTCAACCTGCTG 13634
QY 721 TG 722
Db 13635 AG 13636

RESULT 13
AX333347
LOCUS AX333347 697 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3856 from Patent WO0194629.
ACCESSION AX333347
VERSION AX333347.1 GI:18123981
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 3856 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..697
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 45.7%; Score 581; DB 6; Length 697;
Best Local Similarity 95.1%; Pred. No. 4e-108;
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

QY 620 TGGACACGCTCAGCGTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACG 679
Db 1 TGGACACGCTCAGCGTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACG 60
QY 680 ACAATAACGAGACGGGTACATTCACCGGTCAACCTGACGTGGCGCTTATGTGGCGC 739
Db 61 ACAATAACGAGACGGGTACATTCACCGGTCAACCTGACGTGGCGCTTATGTGGCGC 120
QY 740 GGAACCCGAGAGTGACCTGAAAGAGTGTGACCGCGAGCGCGCTTATGTGGAACCAACG 799
Db 121 GGAACCCGAGAGTGACCTGAAAGAGTGTGACCGCGAGCGCGCTTATGTGGAACCAACG 180
QY 800 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGCTCCCGGGGGA-GCGGGCCCG 858
Db 181 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGCTCCCGGGGGANGCNCNT 240
QY 859 GGGAAAGCGACCCCTGCCCTCAGTGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 918
Db 241 GGGAAAGCGACCCCTGCCCTCAGTGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 300
QY 919 TCTCTGTCCACCCCGCGAGAACACTTTACACGAGGAGATTCGTTTCCAAACCCAGA 978
Db 301 TCTCTGTCCACCCCGCGAGAACACTTTACACGAGGAGATTCGTTTCCAAACCCAGA 360

```

```

QY 979 GGAGATCAATGTACTTACAAAGATCCCATCTATTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 1038
Db 361 GGAGATCAATGTACTTACAAAGATCCCATCTATTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTT 420
QY 1039 ATGACTCTGCAAGCCTTGCTGGTCCAAAGTCAATATGTAATTAATAATATAATAATATAATAATAGAT 1098
Db 421 ATGACTCTGCAAGCCTTGCTGGTCCAAAGTCAATATGTAATTAATAATATAATAATATAATAATAGAT 480
QY 1099 AAGAGCCTATCAATGTATCTTTTGTACAATATGTGTGTAATAATGTAGATCATAGATAGCT 1158
Db 481 AAGAGCCTATCAATGTATCTTTTGTACAATATGTGTGTAATAATGTAGATCATAGATAGCT 540
QY 1159 GACTTTCAGAGTCACATTTATAAGTAATTCACCTTAAGATATATATTTTTCACAACAA 1218
Db 541 GACTTTCAGAGTCACATTTATAAGTAATTCACCTTAAGATATATATTTTTCACAACA--- 597
QY 1219 GTTTTGTACTTTTGAATAAATCTTTCTTTATATTTGCTTAAATAAATAAATAAATAAATAA 1271
Db 598 -AGTTTGCACCTTTTGAATAAATCTTTCTTTATATGCTTAAATAAATAAATAAATAAATAA 649

RESULT 14
AX3333763
LOCUS AX3333763 697 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 4272 from Patent WO0194629.
ACCESSION AX3333763
VERSION AX3333763.1 GI:18124482
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 4272 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..697
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 45.7%; Score 581; DB 6; Length 697;
Best Local Similarity 95.1%; Pred. No. 4e-108;
Matches 621; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

QY 620 TGGACACGCTCAGCGTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACG 679
Db 1 TGGACACGCTCAGCGTGGCGTCCAGCTACATCGCCCACTTGAGGCAGATCCTGGCTAACG 60
QY 680 ACAATAACGAGACGGGTACATTCACCGGTCAACCTGACGTGGCGCTTATGTGGCGC 739
Db 61 ACAATAACGAGACGGGTACATTCACCGGTCAACCTGACGTGGCGCTTATGTGGCGC 120
QY 740 GGAACCCGAGAGTGACCTGAAAGAGTGTGACCGCGAGCGCGCTTATGTGGAACCAACG 799
Db 121 GGAACCCGAGAGTGACCTGAAAGAGTGTGACCGCGAGCGCGCTTATGTGGAACCAACG 180
QY 800 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGCTCCCGGGGGA-GCGGGCCCG 858
Db 181 CGTCTGACCTTGGAGTGGAGTCTGGGAAAGCGCGCTCCCGGGGGANGCNCNT 240
QY 859 GGGAAAGCGACCCCTGCCCTCAGTGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 918
Db 241 GGGAAAGCGACCCCTGCCCTCAGTGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 300
QY 919 TCTCTGTCCACCCCGCGAGAACACTTTACACGAGGAGATTCGTTTCCAAACCCAGA 978

```

Db 301 TCTCTGCCACCCCGGAGAACACTTTTAAACGACGAGGAGATTGTTTCCAAACCCAGA 360  
QY 979 GGAGATCAATTTGACTTACAAAGATTCCCATCTATTATTAACTTTATTAACTTCTACCGTGA 1038  
Db 361 GGAGATCAATTTGACTTACAAAGATTCCCATCTATTATTAACTTTATTAACTTCTACCGTGA 420  
QY 1039 ATGACTCTGCAAGCCTTGCTGGTCCAAAGTGCAATATGTAATATATAAATATAAATAGAT 1098  
Db 421 ATGACTCTGCAAGCCTTGCTGGTCCAAAGTGCAATATGTAATATATAAATATAAATAGAT 480  
QY 1099 AAGAGCCTATCAATGATCTTTTGTACAATATCTTTTAAATGTAGATCATAGGATAGCT 1158  
Db 481 AAGAGCCTATCAATGATCTTTTGTACAATATCTTTTAAATGTAGATCATAGGATAGCT 540  
QY 1159 GACTTTGACAGTCACATTTATATAAGTAATTCATCTTAAAGATATATATTTTTCAAACAA 1218  
Db 541 GACTTTGACAGTCACATTTATATAAGTAATTCATCTTAAAGATATATATTTTTCACACA --- 597  
QY 1219 GTTTTCTGCTACTTTGAAATTAATCTTTCTTTATATGCTTAAATATAAATAAATA 1271  
Db 598 -AGTTGCACTTTTGAATTAACCTTCTTTATATGCTTAAATATAAATAAATAAATA 649

## RESULT 15

CR450293

## LOCUS

DEFINITION

Homo sapiens full open reading frame cDNA clone RZPD0834B111D for  
gene TCF21, transcription factor 21; complete cds; without  
stopcodon.

ACCESSION

CR450293

VERSION

1 (bases 1 to 537)

KEYWORDS

Full ORF shuttle clone, Gateway (TM), complete cds.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 537)

Ebert, L., Schick, M., Neubert, P., Schatten, R., Henze, S. and Korn, B.

Cloning of human full open reading frames in Gateway (TM) system

entry vector (pDONR201)

Unpublished

2 (bases 1 to 537)

Ebert, L., Schick, M., Neubert, P., Schatten, R., Henze, S. and Korn, B.

Direct Submission

Submitted (18-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer

Genomforschung GmbH, Im Neuenheimer Feld 580, D-69120 Heidelberg,

Germany

RZPD; RZPD0834B111D, ORFNo 76

www.rzpd.de/cgi-bin/products/cl.cgi?ClineID=RZPD0834B111D RZPDLIB;

Human Full ORF Clones Gateway (TM) - RZPD (kan-resist.) RZPD LIB No.

834

www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834

www.rzpd.de/products/orfclones/

Contact: Ina Roels

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH,

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 101

Fax: +49 30 32639 111

www.rzpd.de

This clone is available from RZPD;

contact RZPD (customer.service@rzpd.de) for further information.

This CDS clone is a part of a collection of human full length

expression clones generated by RZPD.

This CDS has been inserted into pDONR201 via a BP Clonase (TM)

reaction. Additional sequence has been added in front of the start

codon (ATG): att...AAAAA GGT GGC ACC CCT GGT CCA GGT (ATG)

After the last codon additional sequence has been added: CCA GGC

CCA GGC GGC G in front of the 3' att site (AC CCA GCT TTC TT).

Compared to the reference sequence NM\_003206 we did not find any

amino acid exchanges.

Clone distribution: http://www.rzpd.de/products/orfclones/.

Location/Qualifiers

source 1. .537  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="RZPD0834B111D"  
/clone\_lib="Human Full ORF Clones Gateway (TM) - RZPD"  
/lab\_host="DH10B"  
/note="vector: pDONR201, Site\_1: attP1; Site\_2: attP2"  
1. .537  
/gene="TCF21"  
1. .537  
/gene="TCF21"  
/codon\_start=1  
/protein\_id="CAG29289.1"  
/db\_xref="GI:47496533"  
/translation="MSTGSLSDVEDLQEVEMLECDGLKMDSNKEFVTSNTESSNC  
ENSGPQKRGGLGKRRKAPTKSPKPSGVSGQVQRNANARERARMRLVLSKAFSL  
KTLPLWPVPDPTKLSKLDLTLASSYIAHLRQILANDKYENGYIHPVNLTPFVWAGKP  
ESDLKEVVVTSRLCGTTAS"

## ORIGIN

Query Match 42.2%; Score 537; DB 9; Length 537;  
Best Local Similarity 100.0%; Pred. No. 3.6e-99;  
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 268 ATGTCCACCGGCTCCCTCAGCGATGTGAGACCTTCAAGAGGTGGAGATGTTGGAATGT 327  
Db 1 ATGTCCACCGGCTCCCTCAGCGATGTGAGACCTTCAAGAGGTGGAGATGTTGGAATGT 60  
QY 328 GACGGGTTGAAATGGATTGGAACCAAGAAATTTGTGACTTCCAAACGAGACACCGAGAG 387  
Db 61 GACGGGTTGAAATGGATTGGAACCAAGAAATTTGTGACTTCCAAACGAGACACCGAGAG 120  
QY 388 AGCTTCAACTCGAGATGGTCTCCCGAGAGGCGCGCGCGCTGGGCAAGAGAGG 447  
Db 121 AGCTTCAACTCGAGATGGTCTCCCGAGAGGCGCGCGCGCTGGGCAAGAGAGG 180  
QY 448 AAGCGCGCCCAACCAAGAGAGCCCTCAGCGGGGTCAAGCCAGAGAGGGAAGCAGGTCCAG 507  
Db 191 AAGCGCGCCCAACCAAGAGAGCCCTCAGCGGGGTCAAGCCAGAGAGGGAAGCAGGTCCAG 240  
QY 508 CGCAACGCGCGCAACCGCGAGAGCGGGCCCGCATGCGAGTGTGAGCAAGGCTTCTCC 567  
Db 241 CGCAACGCGCGCAACCGCGAGAGCGGGCCCGCATGCGAGTGTGAGCAAGGCTTCTCC 300  
QY 568 AGACTCAAGACCAACCTGCGCTGGTGCCTCCCGCCACACCAAGCTCTCCAACTGGACACG 627  
Db 301 AGACTCAAGACCAACCTGCGCTGGTGCCTCCCGCCACACCAAGCTCTCCAACTGGACACG 360  
QY 628 CTCAGGCTGGCGTCCAGTACATCGCCCATCTTGAAGGAGATCTCGGTAAACGACAAATAC 687  
Db 361 CTCAGGCTGGCGTCCAGTACATCGCCCATCTTGAAGGAGATCTCGGTAAACGACAAATAC 420  
QY 688 GAGAACGGGTACATTCACCGGTCACCTGAGTGGCGCTTTATGTTGGCGGGAACCC 747  
Db 421 GAGAACGGGTACATTCACCGGTCACCTGAGTGGCGCTTTATGTTGGCGGGAACCC 480  
QY 748 GAGAGTGACCTGAAAAGAGTGGTGCACCGGAGCGCTTATGTGGAACACACCGCGTCC 804  
Db 481 GAGAGTGACCTGAAAAGAGTGGTGCACCGGAGCGCTTATGTGGAACACACCGCGTCC 537

Search completed: December 16, 2004, 19:03:06  
Job time : 5531 secs

**This Page Blank (uspto)**